Human Factors in Robots, Drones and Unmanned Systems, Vol. 138, 2024, 11–21
https://doi.org/10.54941/ahfe1005004

Navigating the Seas of Automation: Human-Informed Synthetic Data Augmentation for Enhanced Maritime Object Detection

Amin Majd, Mehdi Asadi, Juha Kalliovaara, Tero Jokela, and Jarkko Paavola

Turku University of Applied Sciences, Turku, Finland

ABSTRACT

In this work, we propose a novel approach for object detection in marine environments using YOLOv7 (Wang et al., 2022) and Generative Adversarial Networks (GANs) for synthetic data augmentation based on the opinions of a maritime expert. Our proposed method combines the power of YOLOv7 for object detection and utilizes GANs to generate realistic and diverse synthetic data and human information regarding the missing data in our dataset, thereby increasing the size and variability of the training dataset. Experimental results demonstrate the effectiveness of our approach in detecting various marine objects, including motor boat, sailing boat and seamark, in both Simulated weather conditions and real-world environments. Our findings suggest that combining YOLOv7 with GAN-based synthetic data augmentation provides a powerful tool for object detection in marine environments, paving the way for further research and practical applications in fields such as marine biology, oceanography, and maritime surveillance.

Keywords: Human factor, Autonomous boat, Maritime environment, GAN-network

INTRODUCTION

Digitalization and increased autonomy in transportation have the potential to create sustainable, safer, and more efficient service chains, contributing to a better quality of life and global prosperity. Key technologies, including AI, sensor fusion, and deep learning, are already available for autonomous vessels. However, the challenge lies in effectively integrating these technologies, particularly in the complex and dynamic maritime environment.

The demand for autonomous maritime systems has driven the integration of machine learning to enhance intelligence, particularly in object detection with computer vision. This task faces complexities due to factors such as lighting, weather conditions, and waves. However, ensuring the accuracy and trustworthiness of machine learning algorithms poses a significant challenge, primarily related to acquiring a well-prepared dataset. Creating a detailed dataset covering diverse scenarios proves difficult, time-consuming, and costly across various research areas. Data scarcity in maritime settings hampers progress, given the intricate and expensive nature of data collection and labelling. Additionally, the relatively new concept of autonomy in this domain limits the availability of relevant datasets, compounded by challenges posed by diverse weather conditions during data collection.

In 2022, our aim was to build a comprehensive image dataset in Finland's maritime domain, consisting of 120,216 RGB annotated images. Evaluation by a maritime expert revealed a lack of diversity in weather conditions within our dataset, prompting the need to incorporate human opinions.

To overcome data scarcity, especially in varying weather conditions, we propose a novel approach for maritime object detection. Our method employs human-informed synthetic data augmentation using Generative Adversarial Networks (GANs), implemented through 4Sessions-Net (4S-Net). This innovative strategy positively impacts labelled data and addresses challenges related to dataset imbalance and insufficiency.

Synthetic data generation using GAN networks, such as 4S-Net, is a cutting-edge solution to overcome these limitations. This paper introduces 4S-Net, which augments labelled data, positively impacting results. However, the synthetic data's complexity may not match real-world scenarios, necessitating model evaluation with real data.

The dataset, collected in the complex Finnish archipelago, was accurately labelled and extended with synthetic data representing different weather conditions. Comparative analysis involving three CNNs on the original and new datasets, including GAN-generated data, reveals superior accuracy in models trained on the new dataset.

In summary, while digitalization and autonomy offer promise, data scarcity and environmental challenges in maritime settings hinder progress, requiring a high level of understanding and contribution from domain experts. Synthetic data generation through GAN networks based on expert opinion, as demonstrated with 4S-Net, is a key solution resulting in improved model accuracy. This approach not only addresses the limitations of real-world data collection but also contributes to advancing the application of machine learning in maritime autonomy. The results demonstrate significant improvements in accuracy and reliability while simultaneously reducing the cost and time of data collection through the incorporation of expert opinions in dataset creation.

MARITIME AUTONOMY: COMPUTER VISION AND SYNTHETIC DATA

The maritime environment presents a myriad of challenges that demand innovative solutions to ensure safe and efficient operations, particularly in the realm of autonomous boats. The integration of computer vision as a key component in maritime navigation systems has emerged as a transformative advancement, redefining the way we perceive and manage maritime activities. This paper delves into the paramount importance of computer vision in autonomous boats within the maritime domain and explores the role of synthetic data, generated through Generative Adversarial Networks (GANs), in improving the performance and robustness of these systems, especially under varying weather conditions (Goodfellow et al., 2014).

Enhanced Situational Awareness

One of the foremost advantages of computer vision in maritime autonomy is its ability to provide real-time, 360-degree situational awareness. Through the utilization of cameras and sensors, autonomous boats can identify and track objects, detect navigational hazards, and monitor the vessel's surroundings with unparalleled precision. This heightened awareness significantly reduces the risk of collisions, groundings, and other maritime accidents, ultimately enhancing safety in the maritime environment.

Efficient Navigation

Computer vision systems empower autonomous boats with the capability to navigate through complex and dynamic maritime environments. These systems can identify navigational markers, track lanes, and assess the traffic density, thereby enabling boats to optimize routes, make real-time course adjustments, and avoid congested areas, leading to improved efficiency and reduced fuel consumption.

Autonomous Collision Avoidance

The integration of computer vision enables autonomous boats to make split-second decisions when encountering other vessels, obstacles, or even unpredictable elements like debris or wildlife (Zhang et al., 2021). This autonomous collision avoidance functionality significantly reduces the risk of accidents and collisions, making maritime transportation safer and more reliable.

Reduced Human Intervention

Computer vision systems lessen the dependence on human operators, which is crucial for prolonged maritime missions or operations in remote and hazardous environments (Molina-Molina et al., 2021). Autonomous boats can function effectively with minimal human oversight, reducing operational costs and increasing operational flexibility.

Leveraging Synthetic Data With GANs for Enhanced Performance

Despite the evident advantages of computer vision in maritime autonomy, challenges persist, particularly in the context of varying weather conditions. Traditional computer vision models often struggle when faced with adverse weather, such as heavy rain, fog, or low light. Herein lies the significance of synthetic data generation through GANs (Goodfellow et al., 2014).

Data Diversity

GANs can create synthetic data that simulates a wide range of maritime weather conditions, from clear skies to stormy seas (Becktor et al., 2022). This diversity in synthetic data helps train computer vision models to adapt to different environmental scenarios, ensuring their robustness under adverse conditions.

Continuous Learning

By continually exposing computer vision systems to synthetic data, these systems can learn and adapt over time. This dynamic learning process allows autonomous boats to improve their performance gradually, even in challenging maritime environments.

Cost-Efficiency

Acquiring real-world data that spans various weather conditions can be costly and time-consuming. Synthetic data generation through GANs offers a cost-effective alternative, enabling researchers and developers to amass extensive datasets without the need for expensive field trials.

Testing and Validation

Synthetic data also facilitates rigorous testing and validation of computer vision algorithms under controlled conditions (Lin et al., 2023). This ensures that the systems can reliably handle real-world scenarios, thus bolstering their readiness for deployment in autonomous boats operating in diverse maritime environments.

In short, computer vision is an indispensable component in enhancing maritime autonomy, ensuring safer and more efficient operations (Qiao et al., 2021). Synthetic data, generated through GANs, plays a pivotal role in addressing the challenges posed by varying weather conditions, equipping autonomous boats with the adaptability needed to navigate the unpredictable nature of the maritime environment effectively. As the maritime industry continues to embrace autonomy, the fusion of computer vision and synthetic data promises to revolutionize maritime navigation and usher in a new era of safer, more reliable, and efficient maritime transportation.

RELATED WORKS

In this section, we discuss about the related works based on two main concepts: computer vision and GAN networks.

Computer Vision

Generic object detection aims to identify and classify objects within an image, and label them with rectangular bounding boxes indicating the confidence of their existence. There are two main categories of generic object detection methods: those that follow a traditional object detection pipeline, generating region proposals and then classifying each proposal into different object categories, and those that view object detection as a regression or classification problem, using a unified framework to achieve final results (categories and locations) directly.

The region proposal based methods mainly include R-CNN (Girshick, 2014), SPP-net (He et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), R-FCN (Dai et al., 2016), FPN (Lin et al., 2017) and Mask R-CNN (He et al., 2017), some of which are correlated with each other (e.g. SPP-net modifies R-CNN with a SPP layer). The regression/classification based methods mainly includes MultiBox (Chen et al., 2015), AttentionNet (Xu et al., 2015), G-CNN (Zhang et al., 2017), YOLO (Redmon and Farhadi, 2016), SSD (Liu et al., 2016), DSSD (Liu et al., 2017) and DSOD (Liu et al., 2017), which are widely used as marine ship detectors. In general, the speed of the single-stage method is obviously faster than that of the dual-stage method, but the accuracy is close to or even better than that of the latter.

GAN Networks and Synthetic Data

Generative Adversarial Networks (GANs) are a class of deep learning models designed to generate synthetic data, particularly in the context of image datasets. GANs consist of two neural networks, a generator and a discriminator, engaged in a competitive game.

The generator creates fake data samples, aiming to mimic the distribution of real data. Meanwhile, the discriminator evaluates samples, attempting to distinguish real from fake. Through adversarial training, the generator continually improves, making its generated data more realistic.

This framework leverages a minimax game theory approach where the generator aims to minimize the likelihood of being caught by the discriminator, while the discriminator seeks to maximize its detection accuracy. This dynamic equilibrium pushes the generator to generate data that is increasingly indistinguishable from real data.

GANs find extensive use in synthetic data generation for image datasets, as they can create realistic images that are useful for tasks like image synthesis, style transfer, and data augmentation. Their capacity to capture intricate patterns in data has made them pivotal in various computer vision and image processing applications.

PROPOSED WORK

Our workflow consists of three stages: 1) exploring the current datasets by an human expert to define the type of missing data based on different weather conditions, 2) the high-resolution daytime translation (HiDT) (Anokhin et al., n.d.) model for the Daytime Translation task and 3) object detection using the YOLO algorithm.

The HiDT

Object detection model is based on an encoder-decoder architecture. The encoder decomposes an image into its style and content, while the decoder generates a new image by combining the content from the original image and the style from the style image. The two components are combined using the AdaIN connection. The overall architecture of the model consists of a content encoder Ec, which maps the initial image to a 3D tensor using convolutional down sampling layers and residual blocks. The style encoder Es is a fully convolutional network that ends with global pooling and a compressing 1x1 convolutional layer. The generator G processes the content tensor with residual blocks containing AdaIN modules and then up samples it. To create a realistic daytime landscape image, the model should preserve the fine details of the original image. To achieve this, the encoder-decoder architecture is enhanced with skip connections between the down sampling part of the encoder Ec and the up sampling part of the generator G. Regular skip connections would introduce the style of the initial input into the output, so an additional convolutional block with AdaIN is introduced to the skip connections. This allows the model to preserve the fine details of the original image while still generating a realistic daytime landscape image.

YOLOV7

YOLOV7 is a highly efficient and accurate object detection algorithm that outperforms other state-of-the-art models by reducing the number of parameters and computational costs. It uses a faster and more robust network architecture that integrates features more precisely, has a more stable loss function, and is trained more efficiently. YOLOv7 can be trained on small datasets more quickly and with less expensive computational hardware. The algorithm is a single-stage detector that performs object classification and detection simultaneously by looking at the input image or video once. It has three important parts in its architecture: the backbone, neck, and head. The backbone extracts features from the input images, the neck generates feature pyramids, and the head performs the final detection as an output. YOLOv7 introduces several architectural changes, such as compound scaling, the extended efficient layer aggregation network (EELAN), a bag of freebies with planned and reparametrized convolution, coarseness for auxiliary loss, and fineness for lead loss.

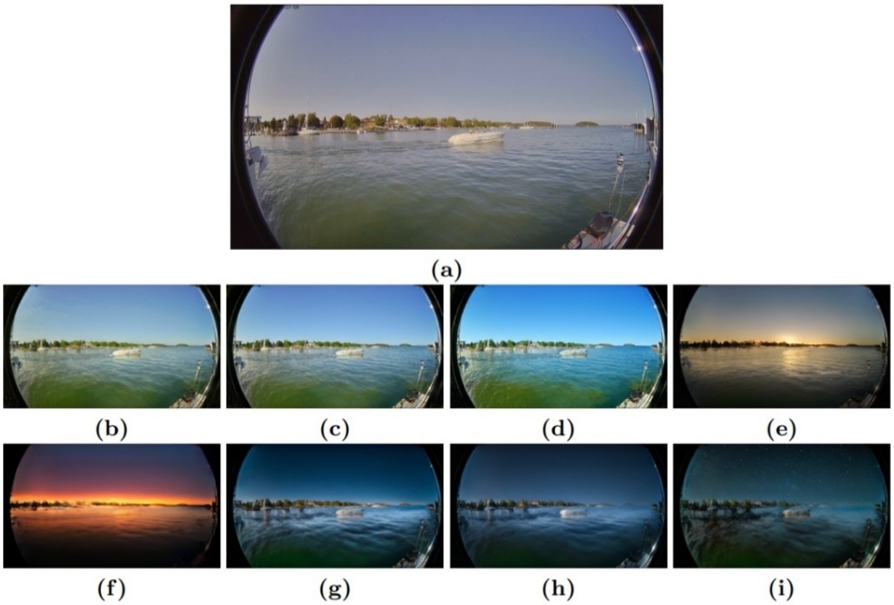


Figure 1: This is some synthetic images; side by side, a) original image b, c, d) three day images e, f) two sunset images and g, h, i) three night images.

MATERIAL AND METHODS

The sensor used for capturing data was the Hikvision DS-2CD2T45G0P, a wide agle camera. Our object detection dataset contains 298 images, of which 199 are in the training set, 49 in the validation set, and 50 in the test set. The task is to detect object categories such as motor boat, sailing boat and seamark. Ground-truth bounding boxes are available, and the evaluation protocol is based on standard mean average precision. The data includes 7,205 labelled objects, and upon examining Figure 1, it is evident that there is a significant bias towards the motor boat class. This is not surprising, as motor boats are the most common surface vehicle in maritime environment. However, this bias will impact the overall results, and it is important to take this into account when evaluating the models. To address this issue, the performance of the model on the most prominent classes will be taken into account, such as by averaging the scores of the most common classes.

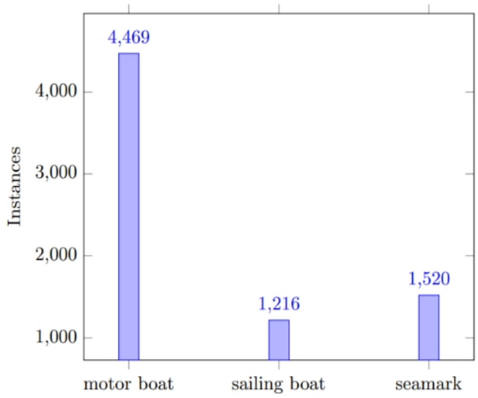


Figure 2: Dataset class instances.

Table 1. The number of data samples.

Dataset	Train Set	Valid. Set	Test Set
Real	199	49	50
Real + Synthetic	3781	49	50

EXPERIMENTAL RESULTS

In this study, two maritime environment experts in the Nordic region examined the existing dataset and identified a need for additional data pertaining to rainy and foggy conditions. Furthermore, our data collection did not include scenarios during nighttime, sunrise, or sunset. In response to expert feedback, we endeavored to generate synthetic data aligning with their insights.

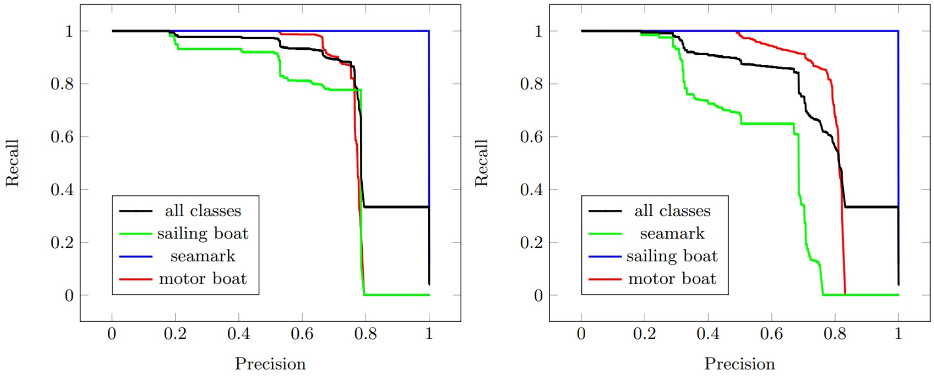
We employ a high-resolution daytime translation (HiDT) model to generate synthetic data and use a pre-trained YOLOv7 model with COCO dataset (Folds et al., 2008) as our object detection model. Our experiments used NVIDIA Tesla V100 GPUs with 32GB memory. The explanations for each step of the analysis will be discussed in the following subsection.

Synthetic Data Augmentation Using HiDT

We utilize the high-resolution daytime translation (HiDT) model to generate synthetic data in different daytimes to augment our dataset. There are several examples of images produced by HiDT that are presented in Figure 1, which demonstrate excellent performance in our dataset. HiDT generates eighteen different weather conditions; we only show eight items in Figure 1. As the experiment of the object detection, we utilized YOLOv7 as an object detection model and analysed YOLO results with and without synthetic data. As demonstrated in Table 2, including synthetic data can significantly improve the mAP of object detection, achieving a mAP of 0.822. Additionally, we tested our models on both real and combined (real and synthetic) datasets, and it was evident in Figure 3 that the models trained on real and synthetic data achieved outstanding results compared to those trained solely on real data.

Table 2. This is mAP@0.5 results of YOLOv7 model.

Train Set	Test Set	mAP@0.5
Real	Real	0.788
Real + Synthetic	Real	0.822
Real	Real + Synthetic	0.518
Real + Synthetic	Real + Synthetic	0.774



(a) (b)

Figure 3: Precision-recall curve for YOLOV7 model a) train on real data + synthetic data, b) train on real data.

DISCUSSION AND CONCLUSION

In this study, we present a practical and innovative solution to the challenge of training a real-time maritime object detector with limited labelled image data. Our approach combines the use of High-Resolution Daytime Translation (HiDT) for data augmentation and the YOLOv7 model for object detection. To tackle the scarcity of labelled data for different weather conditions and times of day, a common obstacle in supervised deep learning, we employ HiDT to generate synthetic images. This not only expands the training dataset but also contributes to the development of a real-time object detector—a valuable tool across various applications. We trained a YOLOV7 model on the augmented dataset, achieving accurate detection with a Mean Average Precision (mAP) of 0.822. This marks a significant improvement over YOLO detection alone, demonstrating the effectiveness of our approach. Our work presents an innovative strategy for enhancing object detection in maritime environments.

REFERENCES

Anokhin, I., Solovev, P., Korzhenkov, D., Kharlamov, A., Khakhulin, T., Silvestrov, A., Nikolenko, S.

Becktor, Jonathan, Schöller, Frederik ET, Boukas, Evangelos, Blanke, Mogens, & Nalpantidis, Lazaros. (2022). "Bolstering Maritime Object Detection with Synthetic Data." *IFAC-PapersOnLine, 55*(31), 64–69. [Elsevier].

Booher, H. R., Minninger, J. (2003) "Human systems integration in army systems acquisition", in: Handbook of human systems integration, Booher, Harold (Ed.). pp. 663–698.

Booher, Harold, ed. (2003). Handbook of human systems integration. New Jersey: Wiley.

Chapanis, A. (1996). Human factors in systems engineering. Wiley Series in Systems Engineering and Management. Andrew Sage, series editor. Hoboken, NJ: Wiley.

Chen, Yolanda, He, Kaiming, & Sun, Jian. (2015). "MultiBox: A fast and accurate object detection method." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(11), 2131–2145. [IEEE].

Dai, Jifeng, He, Kaiming, & Sun, Jian. (2016). "R-FCN: Dense visual object proposals for real-time detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(11), 2131–2145. [IEEE].

Folds, Dennis. Gardner, Douglas and Deal, Steve. (2008). Building Up to the Human Systems Integration Demonstration, INCOSE INSIGHT Volume 11, No. 2.

Friedenthal, S. Moore, A. Steiner, R. (2008) A Practical Guide to SysML: The Systems Modeling Language, Morgan Kaufmann; Elsevier Science.

Girshick, Ross. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(11), 2406–2418. [IEEE].

Girshick, Ross. (2015). "Fast R-CNN." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(11), 2131–2145. [IEEE].

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). "Generative adversarial nets." *Advances in neural information processing systems, 27*.

He, Kaiming, Gkioxari, Georgia, Dollar, Piotr, & Girshick, Ross. (2017). "Mask R-CNN: Object detection and instance segmentation using a mask-based approach." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2131–2145. [IEEE].

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, & Sun, Jian. (2014). "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(11), 2406–2418. [IEEE].

Honour, Eric C. (2006) "A Practical Program of Research to Measure Systems Engineering Return on Investment (SE-ROI)", proceedings of the Sixteenth Annual Symposium of the International Council on Systems Engineering, Orlando, FL.

Lin, Tsung-Yi, Dollar, Piotr, & Girshick, Ross. (2017). "Feature pyramid networks for object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2131–2145. [IEEE].

Lin, Xiaomin, Liu, Cheng, Pattillo, Allen, Yu, Miao, & Aloimonous, Yiannis. (2023). "SeaDroneSim: Simulation of Aerial Images for Detection of Objects Above Water." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 216–223.

Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, & Berg, Alexander C. (2016). "SSD: Single shot multibox detector for real-time object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(11), 2131–2145. [IEEE].

Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, & Berg, Alexander C. (2017). "DSOD: Dual-scale object detection for real-time object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2131–2145. [IEEE].

Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, & Berg, Alexander C. (2017). "DSSD: Dual-scale single shot detector for real-time object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2131–2145. [IEEE].

Meilich, Abe. (2008) INCOSE MBSE Initiative Status of HSI/MBSE Activity (Presentation).

Molina-Molina, J. Carlos, Salhaoui, Marouane, Guerrero-González, Antonio, & Arioua, Mounir. (2021). "Autonomous marine robot based on AI recognition for permanent surveillance in marine protected areas." *Sensors, 21*(8), 2664. [MDPI].

Qiao, Dalei, Liu, Guangzhong, Lv, Taizhi, Li, Wei, & Zhang, Juan. (2021). "Marine vision-based situational awareness using discriminative deep learning: A survey." *Journal of Marine Science and Engineering, 9*(4), 397. [MDPI].

Redmon, Joseph, & Farhadi, Ali. (2016). "You only look once: Unified, real-time object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(11), 2131–2145. [IEEE].

Ren, Shaoqing, He, Kaiming, & Sun, Jian. (2015). "Faster R-CNN: Towards real-time object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(11), 2131–2145. [IEEE].

Taubman, Philip. (June 25, 2008) Top Engineers Shun Military; Concern Grow. The New York Times Website: https://www.nytimes.com/2008/06/25/us/25engineer.html.

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). "{YOLOv7}: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." *arXiv preprint arXiv:2207.02696*.

Xu, Jing, Liu, Ping, Li, Xirong, & Sun, Xing. (2015). "AttentionNet: A bottom-up attention model for visual recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(11), 2131–2145. [IEEE].

Zhang, Jian, Li, Ming-Hsuan, Li, Xirong, & Sun, Xing. (2017). "G-CNN: A graph convolutional neural network for object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2131–2145. [IEEE].

Zhang, X., Wang, C., Jiang, L., An, L., & Yang, R. (2021). "Collision-avoidance navigation systems for Maritime Autonomous Surface Ships: A state of the art survey." *Ocean Engineering, 235*, 109380. [Elsevier].