University of South-Eastern Norway

Development of video processing algorithm (YOLO) in autonomous vessels operations

Candidate name: Behfar Ataei

University of South-Eastern Norway
Faculty of Technology, Natural Sciences and Maritime Sciences

MASTER THESIS

May 2019

Abstract

In recent years, the maritime sector has had a growing interest in the development of autonomous vessels. These new and emerging operational paradigms have gained attention from maritime industry stakeholders through research and development projects, which have subsequently generated a number of both scaled and full-scale vessel prototypes. However, to-date there is a lack of empirical data of real-world autonomous vessels operations. Thus, the majority of research in this domain is limited to conceptual models and feasibility studies.

This thesis focuses on utilizing computer vision as a technology for collision avoidance system in autonomous vessels operations. This study compares the video recordings of a passenger vessel (Ole III) frequently operating in a 120 meters width water crossing using three different methods: (1) observer onboard Ole III (2) manual video observations and (3) computer algorithm. You Only Look Once (YOLO) version 3 is selected as a suitable computer algorithm for video analysis because of its high processing speed and real-time capability. This algorithm is trained by analyzing the video data of passing water traffic and evaluated by comparing it to the observation data manually collected by human observers onboard the vessel and the manual video observations. The main research goal is to test the potential of the YOLO computer algorithm and compare it with the observational data for application in the maritime sector.

By comparing the data processed by the algorithm with the observation data, the algorithm can be evaluated and improved for further applications in autonomous vessels. The higher accuracy of YOLO in detection was attributed to the availability of the data in the training phase. This explains the poor performance of the algorithm when evaluating the data out of the training set. This study applied YOLO to the experimental settings resulting in 95% accuracy in detection, which is within the same range as the benchmarks with similar settings. Although this method shows promise for autonomous vessels applications, further research is required to assess the safety aspects related to the implementation of this technology.

Keywords: Autonomous Vessels, Machine Learning, Object Detection, Convolutional Neural Network, Collision Avoidance, Safety, You Only Look Once

Preface

This thesis is done in Department of Maritime Operations at University of South-Eastern Norway (USN), spring of 2019. This submission is a part of mandatory requirement of the program Master of Science (MSc) degree in Maritime Management, Technical Specialty.

Acknowledgement

I want to thank my family members Parvin, Lohrasb, Yashgin, and Behrouz for all of their Love and Support. It was impossible to overcome the difficulties without their positive energy.

Special thanks to my mentors and supervisors Dr. Steven C. Mallam, Assistant professor Marius Stian Tannum and Dr. Salman Nazir not only for their scientific supervision but also for guiding me through the professional carrier.

Borre, May 2019

Behfar Ataei

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAWA	Advanced Autonomous Waterborne Applications Initiatives
AL	Autonomy Level
ANN	Artificial Neural Networks
ANS	Autonomous Navigation System
CNN	Convolutional Neural Network
AP	Average Precision
COCO	Common Objects in Context
COLREG	Convention on the International Regulations for Preventing Collisions at Sea
DNV-GL	Det Norske Veritas - Germanischer Lloyd
DPA	Norwegian Data Protection Authority
DP	Dynamic Positioning
EC	European Commission
FPS	Frame Per Second
GPS	Global Positioning System
IMO	International Maritime Organization
IOU	Intersection Over Union
LIDAR	Laser Ranging Device
mAP	mean Average Precision
MARKOM	MARitim KOMpetanse
MUNIN	Maritime Unmanned Navigation through Intelligent in Networks
ms	milliseconds
mUSD	million US Dollars
NN	Neural Network
NTNU	Norwegian University of Science and Technology
PASCAL	Pattern Analysis, Statistical Modelling and Computational Learning
RCNN	Region-based Convolutional Neural Network
ReLU	Rectified Linear Unit
ResNet	Residual Network
SAR	Synthetic Aperture Radar
SAR	Search And Rescue
SCC	Shore Control Center
SSD	Single Shot Detector
STPA	System Theoretic Process Analysis
TA	Task Analysis
TEU	Twenty-foot Equivalent Unit
UAS	Unmanned Aerial System
USN	University of South-eastern Norway
v	Version
VOC	Visual Object Classes
YOLO	You Only Look Once

1 Introduction

1.1 Research Background

1.1.1 Autonomous Vessels

Autonomy and automation are the two concepts that have gained significant attention from both academy and industry. From the perspective of a regular user, they may appear the same while it worth mentioning the differences between these two concepts. Automated systems can be operated based on predefined rules given by a supervisor whereas autonomy is achieved when a system can make the best decision based on environmental factors and without the human interference (Parasuraman & Riley, 1997). Autonomy can be seen as a continuous range from no automation as the minimum which the operator decides for all the tasks to the maximum of fully autonomous which the systems decides for every action without any inputs from the supervisor. There are a variety of scales for different autonomy levels in the maritime context. For example, the International Maritime Organization (IMO, 2019) divides the autonomy spectrum into four levels while Lloyds Register has six scales (Lloyd's Register, 2017).

Recent improvements in sensor technologies, computer science, and telecommunication systems facilitated disruptive innovation in the maritime industry, which results in the development of autonomous vessels. The autonomous vessel is "Next generation modular control systems and communications technology will enable wireless monitoring and control functions both on and off board. These will include advanced decision support systems to provide a capability to operate vessels remotely under semi or fully autonomous control" (Waterborne TP, 2011, p. 8). There is a broad range of concerns for the optimum operation of the autonomous vessels such as rules and regulations, insurance, technological reliability, maintenance, operators competency, cyber security (Komianos, 2018). Autonomy research is becoming increasingly popular in order to assess the feasibility of the idea relating to each of the concerns. Industries also are trying to build some prototypes to investigate the practical issues of the developments. One good example of the recent developments in this area is Yara-Birkeland which will start operating in 2020 (Kongsberg Maritime, 2019). Autonomous vessels take advantage of a variety of technologies and devices to assess the environment around them. Some of the examples of these technologies are Cameras, Radar, Laser Ranging Device (LIDAR), Compass and Global Positioning System (GPS) (Elkins, Sellers, & Monach, 2010).

1.1.2 Computer Vision

Artificial Intelligence (AI) is the replication of human intelligence and decision-making ability in computers. Machine Learning is a central part of AI which employs mathematical models to discover the pattern and similarities in the available data in order to predict unexpected future happenings. Machine learning systems do not need to be explicitly programmed for the required task, and that is an advantage. Nowadays, machine learning is becoming more ubiquitous due to four main reasons: i) machine learning often handles complicated tasks more accurately than a human expert ii) an appropriately designed machine learning algorithm does not exhibit bias iii) it is fast to operate and iv) it is economical advantaged (Finlay, 2017).

Computer vision is a subcategory of computer science and more specifically machine learning which focuses on how the computers can understand and interpret the information inside a digital image. Computer vision techniques enable several technologies such as object detection, object classification, scene detection, and face recognition. Although mathematicians laid the groundwork in this area in the late 1960s, computer vision did not appear as a practical tool until recently mainly due to lack of availability of required data, hardware and software. Over the past two decades, significant growth in internet access has led to availability to a massive amount of data. Furthermore, developments in electronics have resulted in an exponential improvement in computation power. Therefore, processing a significant amount of data seems much more feasible than before. The new generations of computer software also made it possible to analyze big data through more optimized algorithms (Goodfellow, 2017).

Conventional vessels utilize human lookout and radars as primary sources for performing the collision avoidance task during sailing while autonomous vessels use sensors for assessment of the environment surrounding the vessel. One of the sensors used in autonomous vessels is the optical cameras which besides computer vision algorithms are converting camera signals to appropriate input for decision-making system onboard the vessel. Amongst the state-of-the-art computer vision algorithms, You Only Look Once (YOLO) algorithm has a higher processing speed in comparison to the other systems. In the current thesis, the focus area is on the appropriateness of YOLO for collision avoidance task in autonomous vessels operations.

1.1.3 Motivation of the Current Thesis

The idea of this research was initiated by Tønsberg municipality focusing on the feasibility of replacing the current passenger vessel (Ole III) with an autonomous one. The research project was defined as “Small Autonomous Ferry” funded by MARKOM (MARitim KOMpetanse) 2020 with the collaboration of University of South-eastern Norway (USN), campus Vestfold and Norwegian University of Science and Technology (NTNU) Ålesund. The main task during the data collection was to perform a risk analysis focusing on assessment of leisure vessels operators behavior in accordance with Convention on the International Regulations for Preventing Collisions at Sea (COLREG) standard (IMO, 2003). Assistant professor Marius Stian Tannum established a secondary objective by video recording of the crossing at the same time interval in order to analyze as a future task.

Since autonomous vessels are at the early phase of development and implementation, there is a lack of empirical operational data in this area. In summer 2018, three datasets (1) video recordings, (2) observation reports and (3) Global Positioning System (GPS) location, heading, throttle, and rudder position of a conventional vessel were collected. Video recordings and observational data are utilized as secondary data sources in this thesis whereas the third data source is analyzed in another master thesis. The mentioned vessel is planned to be replaced with an autonomous vessel in the future. Optical cameras are utilized for collision avoidance task in autonomous vessels. Suitability and accuracy of computer vision algorithm (YOLO v3) for computer vision and related safety concerns are the gaps which are focused on the current thesis.

1.2 Research Objectives

The overall objective of this thesis is to evaluate the potential of computer vision algorithms, in particular, YOLO for autonomous vessel applications such as collision avoidance. This study provides new maritime-related inputs which will potentially contribute to the implementation of emerging autonomous vessels.

The main research questions which are addressed in this thesis are:

RQ 1- Does the developed computer algorithm (YOLO) comply with the observation reports gathered by the crew during watchkeeping?

RQ 2 - To what extent is the machine learning algorithm (YOLO) accurate for application in autonomous vessels operations?

1.3 Thesis Structure

The material presented in this thesis is organized into seven chapters. The summary of the chapters is presented as follows:

Chapter 2: In this chapter, the background of the research is described, and the literature review is presented. Different versions of YOLO are introduced, and the way YOLO v3 performs the detection task is discussed.

Chapter 3: In this chapter, the methodology of the research is presented. The research design is given, and data collection, sampling method, and data analysis are discussed.

Chapter 4: In this chapter results of different methods are shown and compared.

Chapter 5: This chapter interprets the results given in the previous chapter by comparison to the existing literature. The accuracy of YOLO is calculated in this chapter and compared with other methods and benchmarks. Limitations of the current research are discussed here.

Chapter 6: In this chapter, the concluding remarks and recommendation for future research are presented.

Chapter 7: In this chapter, references of the thesis is given.

Appendix A: This section includes the observational form used for manual observer onboard reporting.

2 Background

Technological developments are helping engineering systems to become more and more automated. Due to this idea, autonomous vessels received lots of attention and funding from both academia and industry. Each vessel is consisting of different functionalities such as navigation, propulsion, communication, and supervisory and each of these functionalities can become automated. Based on the automation level in vessels functionalities, different levels of autonomy are assumable. Two of the scales are the International Maritime Organization (see Table 1) which divides the autonomy into four scales (IMO, 2019) while Lloyds Register has six scales of Autonomy Levels (AL) presented in Table 2 from AL1 to AL6 (Lloyd's Register, 2017).

Table 1 International Maritime Organization Autonomy Levels (IMO, 2019)

Description	Definition
Vessel with automated functions and decision making.	There are seafarers onboard to perform the functions and decision making. Some functions might be automated.
Remote control vessel with seafarers onboard.	The vessel is operated form a remote center, but the seafarers are onboard.
Remote control vessel without seafarers onboard.	Vessel operated from a remote center while there are no seafarers onboard.
Fully autonomous	The decision-making system onboard the vessel in fully capable of making decisions and performing the tasks.

Table 2 Lloyds Register Autonomy Levels (Lloyd's Register, 2017)

Autonomy Level	Description	Definition
AL0	Manual	No Autonomous functions. Human is responsible for actions and decision making.
AL1	Onboard Decision Support	All the actions are done by human operator onboard. There are some tools onboard which help for the decision making onboard.
AL2	On and Off-board decision Support	All actions are done by human operator onboard. There are some tools onboard or offboard which can help to improve the decision making onboard.
AL3	Active Human in the loop	Decisions and actions are done by human supervision. Data can be provided onboard or offboard.
AL4	Human in the loop, Operator/supervisory	Decision and actions are done autonomously with human supervision. On highly important tasks human operator has the chance to intervene.
AL5	Fully Autonomous	All decisions and actions are performed by the systems and rarely supervised by human.
AL6	Fully Autonomous	All decisions and actions are performed by the autonomous system and no supervision from the human.

One of the main drives encouraging utilizing this technology is the economy. The research has shown the cost of owning and operating of an autonomous bulk carrier will be 4.3 million US Dollars (mUSD) less in 25 years than a conventional vessel (Kretschmann, Burmeister, & Jahn, 2017). On the other hand, safety should not be sacrificed in order to achieve a higher economic benefit.

2.1 Autonomous Vessels

2.1.1 Developments in Autonomous Vessels

There are a variety of joint projects from industry and academia focusing on the feasibility of autonomous vessels application in the maritime context. The Maritime Unmanned Navigation through Intelligent in Networks (MUNIN) project is funded by European Commission (EC) and it concentrated on assessing the possibility of using autonomous vessels from technical, legal and economic aspects (MUNIN, 2016). The project has shown that the present value of MUNIN autonomous bulk carrier to be seven mUSD higher than the same kind of traditional vessel. Safety in MUNIN project has been divided into two sections of safety and security. Latest accident analysis rates 58% of the maritime accidents are caused by human errors onboard (EMSA, 2018). The MUNIN project has shown that using autonomous vessels can reduce the collision and foundering accident rates to around ten times, and that is mainly due to the elimination of crew and fatigues issues. Another concern which bolded by MUNIN was the cyber security and piracy, and it is shown that these factors can be eliminated by high resilience design while the attractiveness of autonomous vessels for piracy or cyber-attack in unclear. MUNIN project has shown that the legal framework can be modified to allow these vessels to sail while the main barrier is the role of master in law and the way masters responsibilities will be transferred in Shore Control Center (SCC) or any other operational/supervision mechanism (MUNIN, 2016).

Advanced Autonomous Waterborne Applications Initiatives (AAWA) is a project executed by Rolls-Royce with the focus of collaboration between maritime stakeholders in order to investigate the concerns for autonomous vessels operations (Rolls-Royce, 2016). This project focus areas are technology, safety and security, societal and legal acceptance and economy and business models. In the technological section, the project is focused on the application of the latest technological advancements in autonomous vessels. Autonomous vehicles are more developed in the car and aviation industries, and that is the start point. One of the primary outcomes of this project is the Autonomous Navigation System (ANS) architecture which describes how the inputs of sensors can be used for optimized decision making. In the legal section, the status of the autonomous vessel in relation to three groups of jurisdictional rules, technical rules, and private laws are investigated. In the safety and security section, it is focusing on having the same level of safety in comparison to conventional vessels while it might be unachievable in the beginning. At the beginning of autonomous vessels operations, higher contingency levels have to be considered due to uncertainties. In the business section of the project, the main focus is the high-profit potential of this innovation. This technology is a multidisciplinary phenomenon, and its development has to be included in both short term and long-term business plans in order to achieve higher economic benefits.

Another project in this area is the development of an autonomous vessel named ReVolt performed by DNV-GL (Det Norske Veritas - Germanischer Lloyd), the largest classification society worldwide. The target of this project is to replace the road transportation with autonomous vessels in short sea shipping. The prototype is a 60-meter battery driven fully autonomous vessel sailing at 6 knots in maximum 100-knot distances with the capacity of 100 twenty-foot containers. By eliminating the crew, need for superstructure and crew-related facilities will be reduced. It will result in reduced weight, additional cargo capacity and lower operational costs. Utilizing ReVolt in short sea shipping is estimated to decrease the operational expenses about 34 mUSD in comparison to a conventional diesel-driven vessel in 30 year time period which is more than one million dollars per year (DNV GL, 2019).

The latest industrial project is the design and fabrication of zero-emission fully autonomous container vessel named YARA-Birkeland which is the partnership between YARA and Kongsberg companies. YARA-Birkeland is 80-meter length vessel with the carrying capacity of 120 Twenty-foot Equivalent Unit (TEU) between YARA's facilities in Herøya, Brevik and Larvik in Norway. This vessel is planned to reduce 40,000 truck journeys per year, reducing the accidents and pollutions in the supply chain (Kongsberg Maritime, 2019).

2.1.2 Safety

One of the main drivers which enforcing the application of autonomous vessels is the economic factor, and as shown in the projects there is a possibility for economic benefit. Safety of autonomous operations is one of the main concerns in this area and how this concern will resolve is an obstacle. Safety is a concept which defines the design, management, and application of a system will endanger the human life, economy, and environment. Based on the above definition the safety and the risk are highly connected. Risk is a parameter which estimates the effect of hazard to safety while hazards are unpredicted events which might result in danger. Based on the above definitions the Risk (R) can be estimated as a function of the probability of a hazard (P) with its related consequences (C) – see Equation 1 (Kristiansen, 2005).

$$R = P \, . \, C \qquad \textit{Equation 1}$$

The Accident between two vessels is called Collision, and it is one of the main types of accidents in maritime operations. A collision can be divided into three categories of head-on, overtaking and crossing. In Abilio Ramos, Utne, and Mosleh (2019) article different communication scenarios between SCC and autonomous vessel were defined and through Task Analysis (TA), consequences of each failure were estimated.

Thieme, Utne, and Haugen (2018) investigated the application of the available risk models for using in the autonomous vessels area. Nine assessment criteria were selected, and 64 models were reviewed. Ten out of sixty-four models fulfilled six or more criteria while more in-depth investigations showed that none of the models are suitable for direct implementation in risk analysis and assessment of autonomous vessels.

Wróbel, Montewka, and Kujala (2018) are focusing on developing a risk model for the safe operations of the autonomous vessels by using System Theoretic Process Analysis (STPA) approach. On this research, researchers divided the unmanned vessel to the subsystems groups of shore facility, communication systems, within the vessel, interaction with environment and organization environment. The connections between the subsystems have been reviewed, and possible hazards and consequences were assessed qualitatively.

Safety in the transition from conventional to autonomous vessels is one of the main concerns, which is not covered sufficiently in the literature. One research in this area that is trying to fill the gap is performed by Wróbel, Montewka, and Kujala (2017). In this article by assuming the replacement of the conventional vessels with autonomous vessels, different maritime accidents were reviewed, and the possibilities and the consequences of the accidents were re-assessed. The study showed that the frequency of accidents with navigational basis (such as collision and grounding) might reduce while the consequences of non-navigational accidents (such as structural failure and fire) might increase.

2.1.3 Focus Area

An autonomous vessel is the composition of different systems, subsystems, and functionalities. These systems are sensory, navigation, decision making, propulsion, and communication. In Figure 1 the AAWA project architecture of an autonomous vessel is presented. This architecture is built upon Rolls Royce Dynamic Positioning (DP) System (Rolls-Royce, 2016).

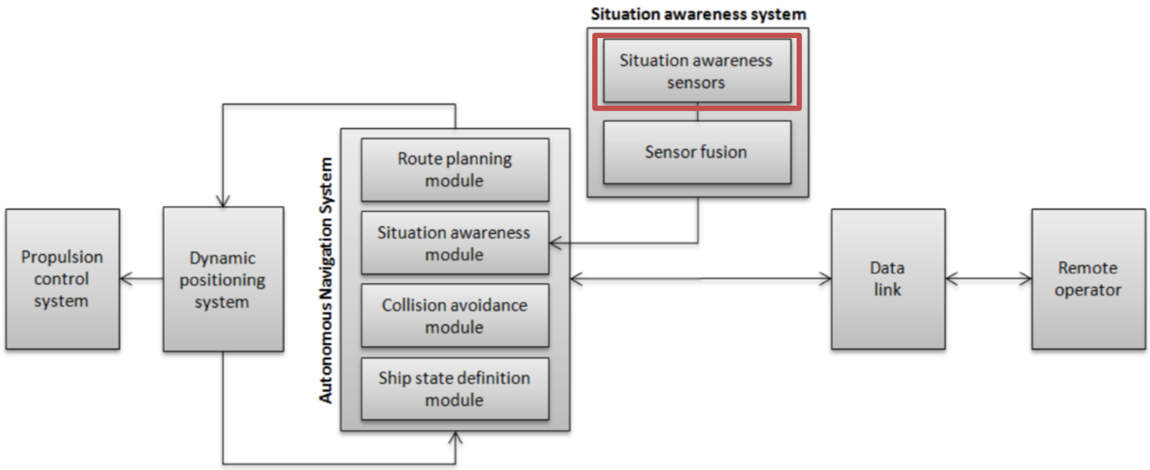

Figure 1 Autonomous Navigation System (ANS) architecture

(Rolls-Royce, 2016, p. 20)

Based on ANS architecture, different sensors are providing the inputs for the autonomous navigation system. Optical cameras are one of the sensors used for collision avoidance task in order to improve safety. Computer vision algorithms such as YOLO can convert the camera signals to appropriate input for the navigation system and perform the detection task. Application of YOLO in autonomous vessels and its accuracy are the main focus areas in the current thesis.

2.2 Neural Network

Neural Network (NN) or Artificial Neural Network (ANN) is a subcategory of Artificial Intelligence (AI) and more specific machine learning. These networks are capable of solving more complicated problems than other machine learning techniques such as decision trees. The capacity of the neural networks can be adjusted by adding or removing the layers and neurons, which is an advantage of these networks (Finlay, 2017).

Neural Networks have different types of feed-forward neural networks, radial basis function networks, multi perception neural networks, convolutional neural networks, recurrent neural networks, and long short-term memory. There are different algorithms used in object detection such as Region-based Convolutional Neural Network (RCNN), Residual Networks (ResNets), Single Shot Detector (SSD) and You Only Look Once (YOLO). All of the mentioned algorithms are using Convolutional Neural Network (CNN) because of grid type nature of inputs (images) (Goodfellow, 2017). The main differences between different technologies are the network architecture, localization (distinguishing the locations of the object in the image) and classification (belonging to a specific group). The different approaches can be compared to each other based on the accuracy and the implementation speed of the algorithm (Goodfellow, 2017).

The idea of RCNN has started by selective search technique. In this method, a small region is selected as an object center, and by grouping backward, the region is expanding, and detection and localization are happening (Uijlings, Sande, Gevers, & Smeulders, 2012). Updated versions of RCNN are using region proposals for performing the detection task (Ren, He, Girshick, & Sun, 2016). Residual networks are made up of residual blocks and mostly used in very deep networks (He, Zhang, Ren, & Sun, 2015). The main advantages of these networks are easier training and higher accuracy. SSD and YOLO are both performing the localization and detection in a single feedforward run. SSD is slower than the YOLO without considerable improvement in accuracy (Redmon & Farhadi, 2018).

2.2.1 Applications of Neural Networks

Performing a literature review, it is recognized that machine learning algorithms are utilized more in the automobile and aviation industries in comparison to the maritime industry. The reason might be more extensive research on autonomous cars and autopilot systems.

The research done by Kim, Hong, Choi, and Kim (2018) is implementing Faster-RCNN (an improved version of RCNN) for the application in the maritime context. The goal of the research is the improvement of detection accuracy by performing Intersection Over Union (IOU) tracking and utilization of Bayesian Fusion.

It is always beneficial to share experiences between industries and even combining them. Research done by Rodin et al. (2018) is combining aviation and maritime industries. The research is performed by utilizing Unmanned Aerial System (UAS) for Search And Rescue

(SAR) operation in the maritime sector. The target is to detect and classify the floating objects from the images taken by the planes using a CNN. The network used in this research achieved 92.5% accuracy on the detection task.

One example of applying YOLO in the maritime industry focuses on vessels detection using YOLO V2 algorithm. Synthetic Aperture Radar (SAR) images are one of the maritime traffic monitoring sources and researchers are trying to apply YOLO v2 for the detection task on them. In this work, the original YOLO V2 and a reduced version of this algorithm were implemented, and the detection accuracy and speed were compared. Both of the networks maintained 90% detection accuracy while reduced YOLO achieved higher processing speed (Yang-Lang et al., 2019).

YOLO is one of the fastest object detection systems which can be employed for real-time detection tasks or video analyses. Limited amount of research is performed on the application of YOLO in the maritime sector, and it is a gap which current thesis is trying to fulfill.

2.2.2 You Only Look Once (YOLO)

When it comes to real-time object detection and video processing, the speed of the approach is a critical factor. In the old object detection systems, the detection and classification were happening in two different steps which will result in extended processing time, and that is why You Only Look Once (YOLO) invented. In the YOLO algorithm, the input is fed forward to the network, and the localization and detection are happening in the single phase which will result in faster processing time. YOLO has the speed of 45 Frame Per Second (FPS) with the accuracy of 63.4 mean Average Precision (mAP) on the PASCAL VOC 2007 Dataset (Redmon J., Divvala, Girshick, & Farhadi, 2016).

Pattern Analysis, Statistical Modelling and Computational Learning (PASCAL) Visual Object Classes (VOC) challenge is an online competition for the object detection, classification and segmentation tasks. In this competition each year the organizers are providing standards datasets for competition and participants are competing to achieve the highest accuracy (PASCAL, 2019).

2.2.3 YOLO 9000 (version 2)

On the second version of YOLO, the main focus area was to improve the accuracy and increase the processing speed at the same time. The updates on YOLO 9000 (version 2) in comparison to original YOLO were the use of batch normalization, high-resolution classifier and using anchor boxes for detection. YOLO version 2 has the mAP of 76.8 at 67 FPS on VOC 2007 dataset (Redmon & Farhadi, 2016).

2.2.4 YOLO Version 3

The main target for developing YOLO v3 was to improve the accuracy of the detections while keeping the processing speed as high as possible. On this version of YOLO, the independent logistic classifiers will be used for class detection, and binary cross entropy loss will be used during training. The feature extractor is changed from Darknet 19 with 19 convolutional layers to the Darknet 53 with 53 layers which is increased the depth of the network. YOLO v3 has 57.9 Average Precision (AP) on 0.5 Intersection Over Union (IOU) in COCO dataset (Redmon & Farhadi, 2018).

Common Objects in Context (COCO) is a competition in Artificial Intelligence (AI) focusing on object detection and labeling. COCO is providing a standard foundation for competition between scientist, entrepreneurs, and researchers working in this sector (COCO, 2019). The accuracy of different detection systems based on COCO dataset is presented in Table 3. On this table numbers, 50 and 75 are IOU percentage and S, M and L are presenting Small, Medium and Large size objects respectively.

Table 3 Accuracy of Different detection systems on COCO dataset

(Redmon & Farhadi, 2018, p. 3)

	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI	Inception-ResNet-v2	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
One-stage methods							
YOLOv2	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD513	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet	ResNeXt-101-FPN	40.8	61.1	44.1	24.1	44.2	51.2
YOLOv3 608x608	**Darknet-53**	**33.0**	**57.9**	**34.4**	**18.3**	**35.4**	**41.9**

Through this document wherever the word "YOLO" is mentioned without its specific version "YOLO v3" is meant.

2.3 Neural Network Architecture

The idea of the neural network comes from neural systems of living creatures. A neural network is consisting of different cells which called neurons and their connections (edges). A neural network is consisting of different layers of neurons. The input layer is the layer which the data feed to the network while the result will be seen on output layers. Between the input layer and the output layer hidden layers are located. They named as hidden layer since the results of these layers are mostly vectors or matrices which does not have a sensible meaning to the user. Each of the neurons contains an activation function which transforms the input of the cell to the output (Di, 2018). A sample of the neural network architecture is presented in Figure 2.

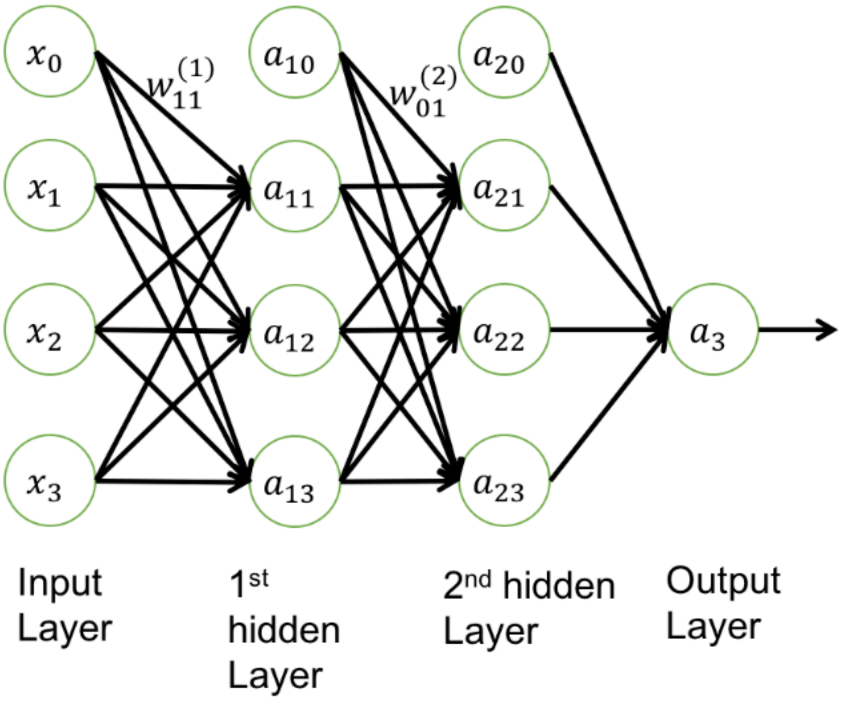

Figure 2 Neural Network Sample Architecture (Di, 2018)

2.4 Convolutional Neural Network

Convolutional Neural Network (CNN) is a specific type of neural network which has an application on the processing of special data with grid type architecture such as time series data or images. As is clear from the name on this architecture the mathematical convolution function has been used for this network. The convolution function can be defined as Equation 2 which the x is the input function, and w is Weight Function, Kernel or Mask (Goodfellow, 2017).

$$s(t) = \int x(a)w(t-a)da \quad \text{Equation 2}$$

The convolution function can also be shown with the asterisk sign (see Equation 3).

$$s(t) = (x * w)(t) \qquad \text{Equation 3}$$

The output of the convolution function is named the feature map. In the machine learning application, the inputs and kernels are multidimensional arrays which will be justified by the training phase. In the case of two-dimensional input such as image, the convolutional function changes to form shown in Equation 4. The reason that the integral has changed to series is due to discrete data. One example of the convolution operation is shown in Figure 3.

$$S(j,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m,j-n) \text{ Equation 4}$$

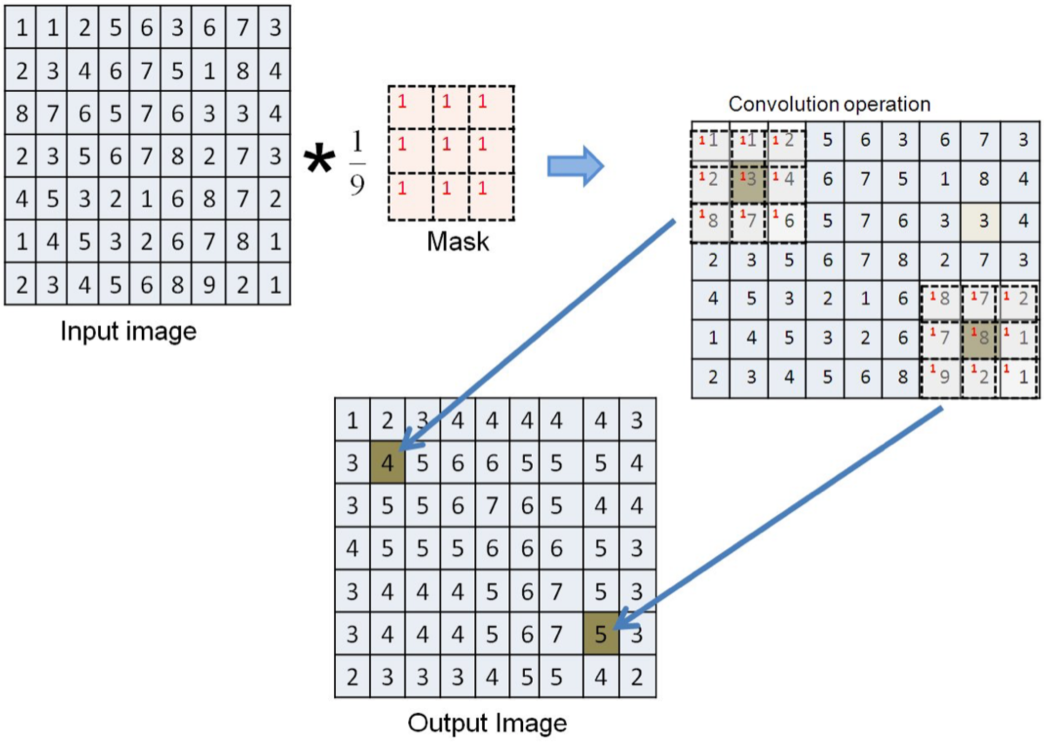

Figure 3 Convolution Operation (Kapur, 2017)

The main applications of Kernels are feature extraction. In earlier layers of networks simpler forms of kernels are being used for simpler detections such as edge detection while other more complicated filters such as face detections are implemented in the deeper layers (Kapur, 2017). An example of a Sobel edge detector is shown in Figure 4.

Figure 4 Example of Sobel Edge Detector (Kapur, 2017)

2.5 You Only Look Once (YOLO) Version 3 Description

The architecture of YOLO is presented in Figure 5 which is consists of convolutional and residual layers. The whole network has 106 layers. Darknet 53 is located at the beginning of the network for feature extraction. Darknet is the name of the developer team, and 53 is the number of convolutional layers in the network (Redmon & Farhadi, 2018). Using stride size in convolutional layers will shrink the feature maps through the network, while filters will increase the volume of the features. The future map size will reduce to a certain degree and then it will expand using upsampling. The detections are taking place in layers 82, 94 and 106.

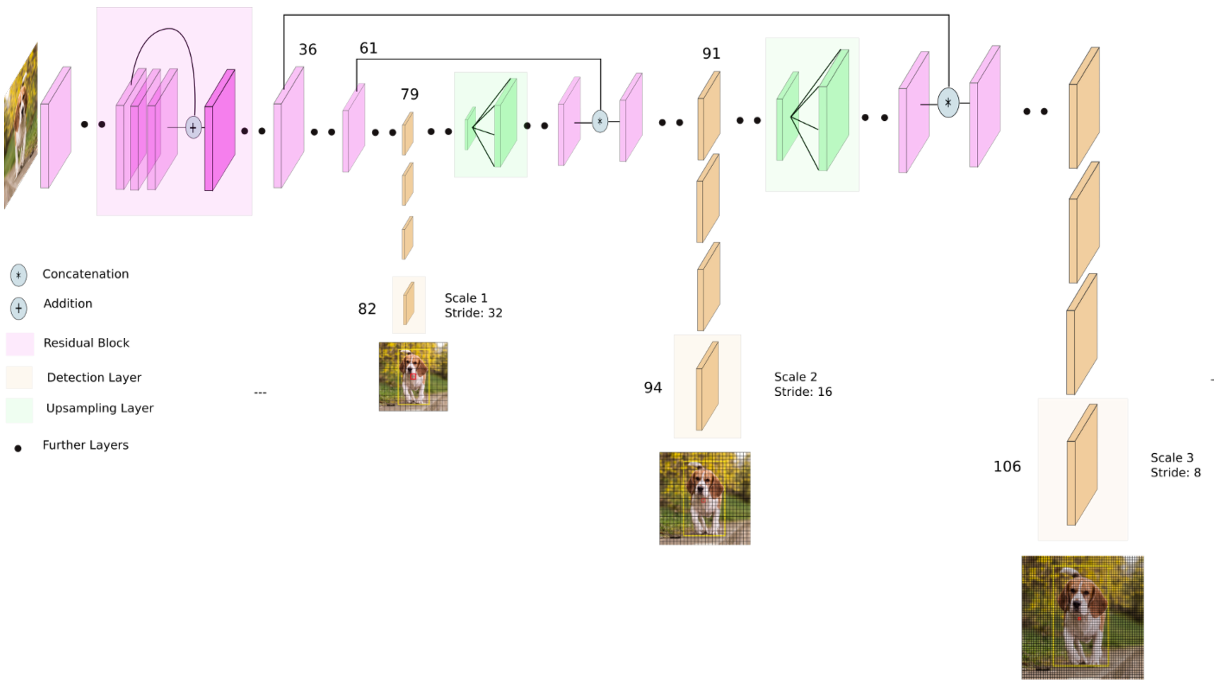

Figure 5 Yolo v3 Network Architecture (Kathuria, 2018)

2.5.1 Activation Function

In Figure 6 a simplified neural network is presented. In the feed-forward process, each of the neurons is receiving an input ($a^{(L-1)}$), they multiply input to the weight ($w^{(L)}$) and then add a bias value ($b^{(L)}$) (see Equation 5). Each of the neurons contains a non-linear function which named as activation function (σ). Utilizing non-linear functions increase the network ability to perform more complicated tasks (Goodfellow, 2017). Neurons will implement the activation function (see Equation 6), and the result will be the output of layer given to the next layer as input.

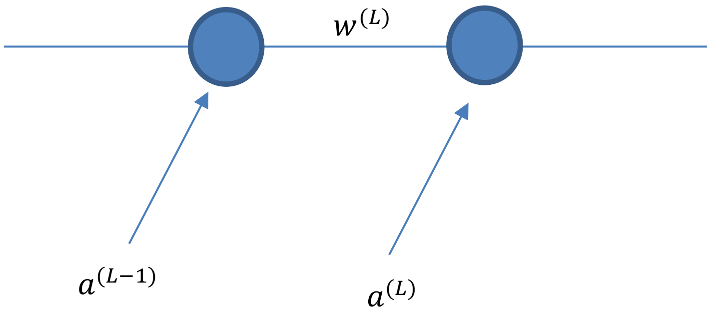

Figure 6 Simplified Neural Network

$$Z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)} \qquad \textit{Equation 5}$$

$$a^{(L)} = \sigma(z^{(L)}) \qquad \textit{Equation 6}$$

There are a variety of activation functions used the neural networks such as tanh, sigmoid, Rectified Linear Unit (ReLU), Leaky ReLU. Each of the activation functions has its advantages and disadvantages. As an example, using tanh has weaknesses such as activation saturation and convergence difficulties. ReLU is faster in convergence but has issues with dead neurons. Leaky ReLU (see Equation 7) claimed to solve dead neurons issue in RELU while it has not proved yet (Goodfellow, 2017). YOLO is using Leaky ReLU as activation function which is shown in Figure 7.

$$f(x) = \begin{cases} x & if\ x > 0 \\ 0.01x & otherwise \end{cases} \qquad \textit{Equation 7}$$

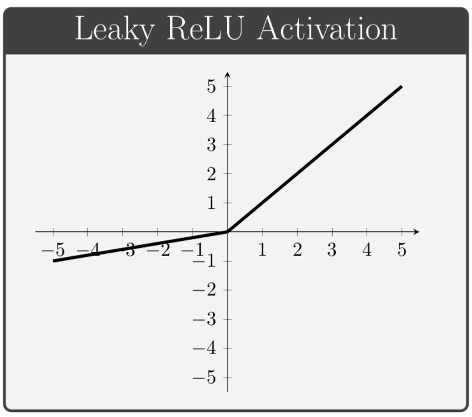

Figure 7 Leaky ReLU Activation (Winovich, 2019)

2.5.2 Residual Block

One of the main improvements in the YOLO v3 in comparison to the previous versions is the use of residual blocks. In the classic neural networks, features are learned one by one and the output of one layer is used as input of the next layer. The residual block is made of a shortcut path as shown by the arrow (see Figure 8). The result of the block can be divided into the old feature (x) and residual (F(x)). In this architecture, the layer inside the block learns what to add to the old feature to produce better output. Using residual blocks facilitate the training process in deeper networks with a higher number of layers (He et al., 2015).

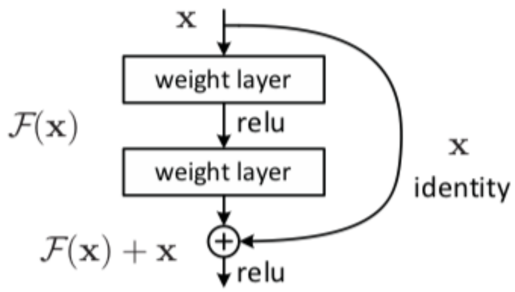

Figure 8 Residual Block (He et al., 2015, p. 2)

2.5.3 Concatenation

During the feed forward phase of the neural network the feature map shrinks and the chance of losing details in the network increases. Concatenation is used in YOLO to reduce the effect. In YOLO the feature map size reduces to a certain degree and then using upsampling it starts to expand. The feature maps with the same size are compared and concatenated before feeding to the detector (Redmon & Farhadi, 2018).

2.5.4 Detection

Previous versions of YOLO were suffering from lower detection accuracy of small objects while YOLO v3 overcome the problem by performing the detection in three different layers and scales (layers number 82, 84 and 106). In layer 82 the feature map has the size of 13x13x255 and detection of the biggest objects are performing as the feature map has the smallest size. In layer 94 the network size is 26x26x255 and detection of medium size objects are happening. In layer 106 the feature map size is 52x52x255, and detection of small objects are performing (Redmon & Farhadi, 2018).

2.5.5 Classification

YOLO v3 is using the logistic regression for the classification task. Using this technique allows the network to perform the multilabel classification. One of the issues for using SoftMax is that the classes should be mutually exclusive while in independent logistic classifiers the concern is solved (Redmon & Farhadi, 2018).

2.5.6 Loss Function

In machine learning algorithms, loss functions are used for assessing the accuracy of the network in the training phase and updating the hyperparameters (weights, biases, and kernels). During the training phase at the end of each loop, the network outputs are compared with the actual values and error is calculated using the loss function (Goodfellow, 2017). The loss functions of YOLO have been presented in Equation 8 (Redmon J. et al., 2016).

Equation 8 (Redmon J. et al., 2016)

$$F(loss) = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

$$+\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} (C_i - \hat{C}_i)^2$$

$$+\lambda_{noob} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} (C_i - \hat{C}_i)^2$$

$$+ \sum_{i=0}^{s^2} 1_{i}^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2$$

The term i defines the grid cell number, while j defines the bounding box number. In the above function, the first term is responsible for locating the centroid of the object, and the second term defines the height and width of the object. The third and fourth terms are defining the confidence of the detection, and the fifth term is responsible for the classification of the objects.

2.5.7 Backpropagation

The main idea in the training phase of the algorithm is to reduce the cost function. At the beginning of training phase kernels, weights and the biases are assigned randomly. In a feed-forward process, the inputs are fed to the algorithm, and the outputs are calculated.

After finishing the forward process by comparing the algorithm output and the actual output the cost will be calculated (see Equation 9). In the training phase of the algorithm, the kernels, weights, and biases of the neural network have to be adjusted, and the backpropagation algorithm is one of the ways to perform this task. The main idea is how the cost function (C) is changing with the change of weight or bias with using partial derivatives such as ∂C/∂w. The partial derivative of the cost function is calculated based on each of the dependent variables. In order to minimize the cost, the direction of the change for each of the variables will be calculated using the chain rule (see Equations 10, 11, 12 and 13). The weights and the biases have to be adjusted to reach the best output and moving toward the global minimum of the cost function. Stochastic gradient descent can be used to facilitate the optimization process by moving toward the value with highest change rate (Goodfellow, 2017).

$$C_0(\dots) = (a^{(L)} - y)^2 \qquad \textit{Equation 9}$$

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial C_0}{\partial a^{(L)}} \qquad \textit{Equation 10}$$

$$\frac{\partial C_0}{\partial a^{(L)}} = 2(a^{(L)} - y) \qquad \textit{Equation 11}$$

$$\frac{\partial a^{(L)}}{\partial z^{(L)}} = \sigma'(z^{(L)}) \qquad \textit{Equation 12}$$

$$\frac{\partial z^{(L)}}{\partial w^{(L)}} = a^{(L-1)} \qquad \textit{Equation 13}$$

2.5.8 Anchor Boxes

Yolo uses predefined boxes sizes for the detection task. Yolo v3 uses nine anchor sizes at three different scales of detection. The three biggest anchors will be used in the first detection layer for the detection of large items in the smallest feature map while the smallest anchors will be used in the last detection layer with biggest feature map to detect small objects. Medium sized anchors are used for detection of objects with medium sizes in middle detection layer. The size of the anchor boxes will be calculated (as shown in Figure 9) based on the training dataset using the K mean clustering (Redmon & Farhadi, 2018).

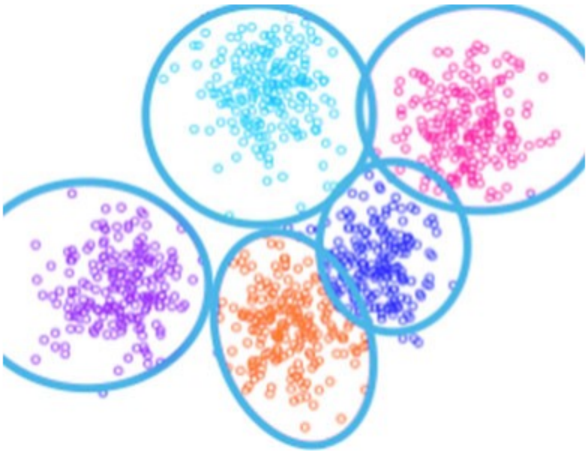

Figure 9 K-mean clustering concept used for Anchor Box sizing (Hui, 2018)

2.5.9 Non-max Suppression

In order to prohibit the multiple detections of the same object non-max suppression is used. In this approach, the bounding box with maximum confidence score is considered as the primary detection. The Intersection Over Union (IOU) is calculated between the primary detection and secondary detection (see Equation 14 and Figure 10) and compared with a threshold. If the IOU is larger than the threshold, it is a repetitive detection, and the object with lower confidence is eliminated while if the IOU is lower than the threshold, they are two different objects (Redmon J. et al., 2016).

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{A \cap B}{A \cup B} \qquad Equation\ 14$$

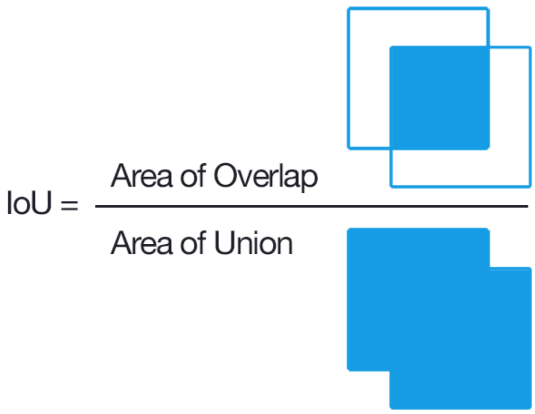

Figure 10 Intersection Over Union (IOU) (Stack Overflow, 2017)

2.6 Fit, Underfit, and Overfit

The main target in machine learning is to have high performance on the unforeseen data which was not used in the training phase. In the training phase loss function is minimized by updating the kernels, model weights, and biases. One of the issues that might arise is, to what extent the parameters in the neural network should update. This issue is called underfitting and overfitting (see Figure 11). Underfitting is the situation where the model is unable to lower the error in the training dataset and has low performance in the training dataset. Overfitting is the situation when parameters in the neural network are updated excessively. It means that the model memorizes the features in the training set and has high accuracy in the training dataset while it is not able to predict accurately on the test dataset. In order to control the underfit or overfit situation, the capacity of the model can be modified. Models with a lower capacity are more likely to be underfitted while models with large capacity can overfit. That is why the model with appropriate capacity has to be used (Goodfellow, 2017).

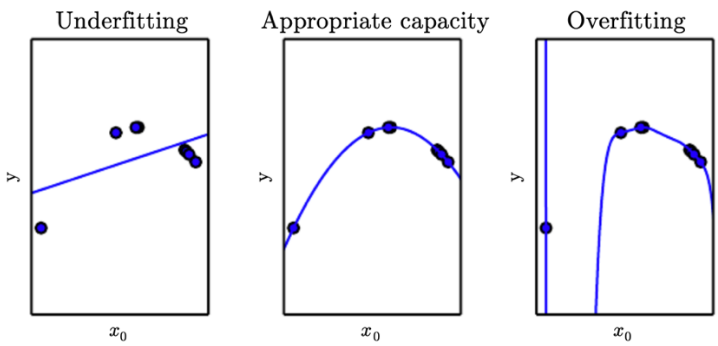

Figure 11 Example of different fit for a polynomial model

(Linear Model, Quadratic Function, Polynomial Degree 9) (Goodfellow, 2017, p. 110)

3 Methodology

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed. Same as the neurons in the brain that can be trained to pass only signals that are useful in achieving the broader goals of the brain, we can train the neurons of a neural network to pass along only useful signals (Di, 2018). The recent development of Neural Network has shown that the Convolutional Neural Network (CNN) is the most suitable type in the computer vision application (Kapur, 2017). In the current thesis, one of the objects is to give the video data as input to a neural network and receive the detected vessels in the video as output.

The video data is divided into three parts of the training, test, and evaluation. By using the training data, the algorithm will learn the desired pattern in the data while with the test data the performance of the algorithm will be assessed. Supervised learning is performed when data are explicitly labeled. In this thesis, one of the objectives is to classify the vessels types, and that is why supervised learning is used (Goodfellow, 2017). After analyzing the evaluation dataset by the computer algorithm, the results can be compared with the manual video observation and observational onboard to be validated.

In the observational area, the traffic pattern is highly complex, and that is due to types of crossings, where the majority of the traffic are leisure vessels. It has seen that in some cases, crossing traffic are not following the International Regulation for Preventing Collision at Sea (COLREG) (IMO, 2003) and that might be because of lack of training of the operators. The inadequate training is increasing the risk of operating for the passenger vessel, and it demands more attention to the accuracy of the data processing systems.

3.1 Research Design

The current thesis is conducted based on Ole III vessel and its specific voyages in Husøy-Husvik route which is considered as a single case study. One of the main drawbacks of having a single case study is the low generalization level. Moreover, the feasibility of the research on other settings is unclear (Frankfort-Nachmias, 2015). The overall research design is presented in Figure 12. It should be mentioned that the research design is not a sequential process and it is cyclical (Frankfort-Nachmias, 2015) and only for the simplicity of visualization it is shown as a linear model. Data collection were performed as the scope of another project before the execution of the research design on the current thesis. Utilizing secondary data facilitates the research process while enforces some limitations (Sharp, Peters, & Howard, 2002). Two types of observational reports and video recordings were collected in the data collection stage. A literature review was performed to identify gaps in the research area and to select the appropriate methodology. The video data is analyzed using the two different methods of YOLO and manual observation. The video data is divided into three parts of training, test, and evaluation. The training data is used for the training phase of YOLO while test data is used to assessing YOLO performance in this stage. On this research, the manual video observation, observational reports from observer onboard Ole III and YOLO are compared, and machine learning algorithm is validated.

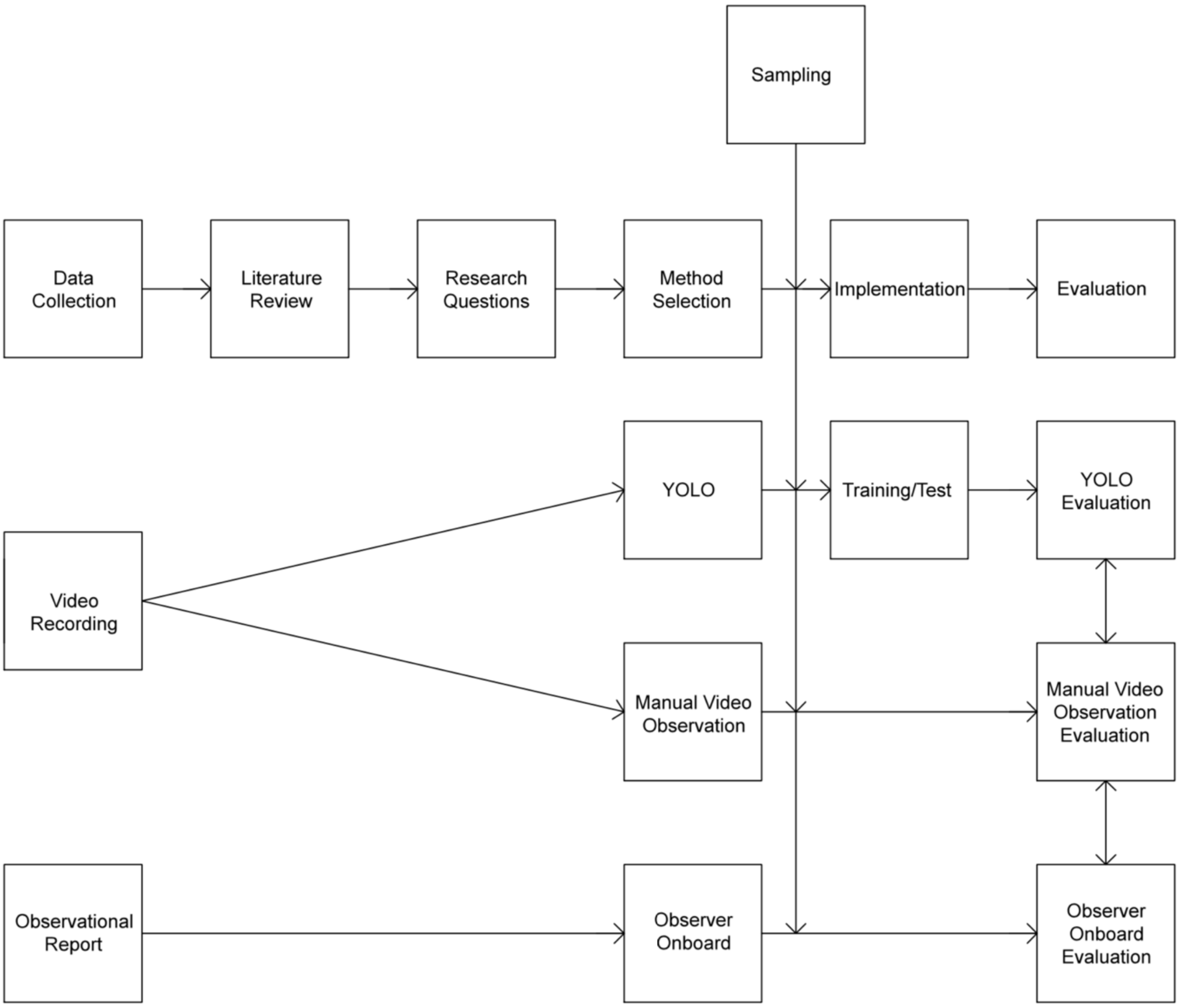

Figure 12 Overview of research design

3.2 Method Selection (YOLO)

The former neural networks performing the object detection task in two stages. In the first stage, the location of the object in the image is detected, and in the second stage, classification is occurring. In the YOLO algorithm, as the name shows, both localization and classification are happening at the same time, and it results in increased speed.

Figure 13 presents the various computer vision algorithm for object detection purposes. The horizontal axis is presenting the processing time for one frame in milliseconds (ms), and the vertical axis is the Average Precision (AP) of the algorithm in COCO dataset (Redmon & Farhadi, 2018). Suitable computer vision algorithm can be selected based on the balance between computational power requirement, speed, and accuracy.

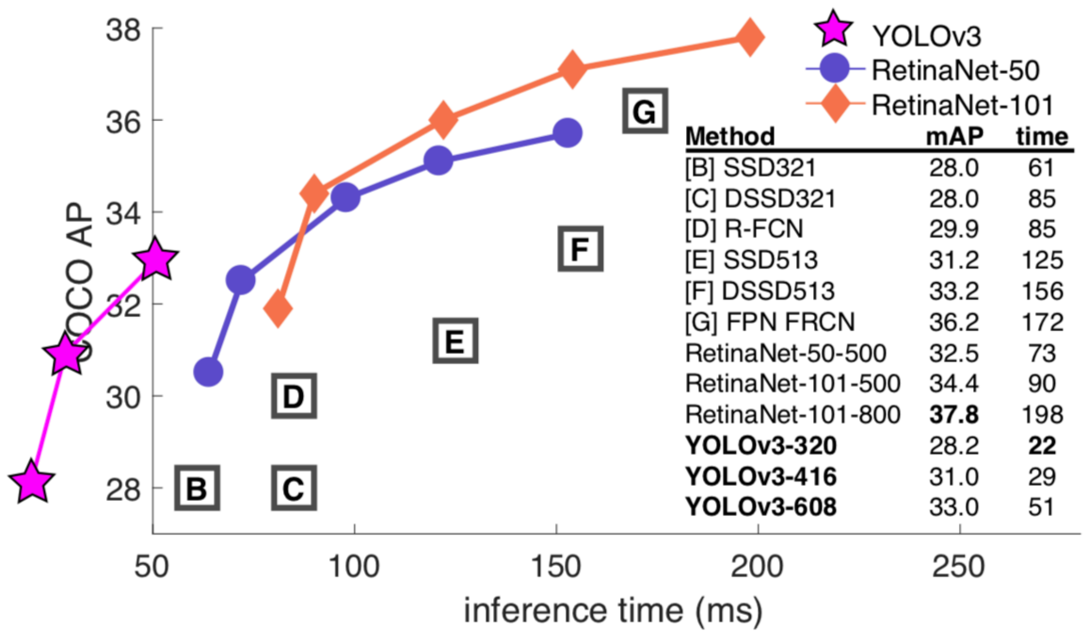

Figure 13 Accuracy/Time Curve for Detection Systems (Redmon & Farhadi, 2018)

Implementing the object detection systems are demanding lots of computational power which is a bottleneck for all phases of training, test, evaluation, and implementation. In our case since the leisure vessels will be evaluated and also the rate of the change in the environment is considerably high. We decided to select the YOLO v3 network as the technique for the object detection task and the target is to evaluate the applicability of this model for the maritime usages. On the current thesis, codes from AlexeyAB is used as the source for YOLO implementation (AlexeyAB, 2019a).

Object detection systems are sensitive to the network size for the detection of small-sized objects. Networks with bigger sizes have a higher chance for the detection of small objects. YOLO can be adjusted for three different sizes of 320x320, 416x416 and 608x608 pixels (Redmon & Farhadi, 2018). In this thesis, the biggest network size (608x608) is used to improve the accuracy of detections.

3.3 Data collection

During summer 2018, two sets of observation and video data were collected from the "Ole III". The vessel is operating in the Husøysund area in Husøy-Husvik route located in Tønsberg municipality in Vestfold county, Norway. The observational data were collected by University of South-eastern Norway (USN) nautical science bachelor students during the sailing from 9:00 until 20:00 (from 4th June to 4th August 2018) while the video data was collected by an automatic video recording system from 7:00 to 21:30 (from 6th June to 27th August 2018).

The following types of data have been collected.

- Two Sets of Video Data
  - Optical Camera
  - Thermal Camera
- Observation Data
  - Time and Date
  - Passengers Data (Number of passengers, bicyclists, demanding special help)
  - Violations of COLREG by passing traffic
  - Captain Decision in case of COLREG violation
  - Crossing Traffic Data in case of Intervening Navigation (Number of vessels and types)

3.3.1 OLE III

The idea of the research was initiated by Tønsberg municipality about the possibility of replacing of current passenger vessel with an autonomous one. The crossing is about 120 meters and the vessel OLE III is transporting the passengers between the small quays located in each side (see Figure 14). Ole III has the maximum capacity of 11 passengers, and it can transport bicyclists and passengers demanding special help.

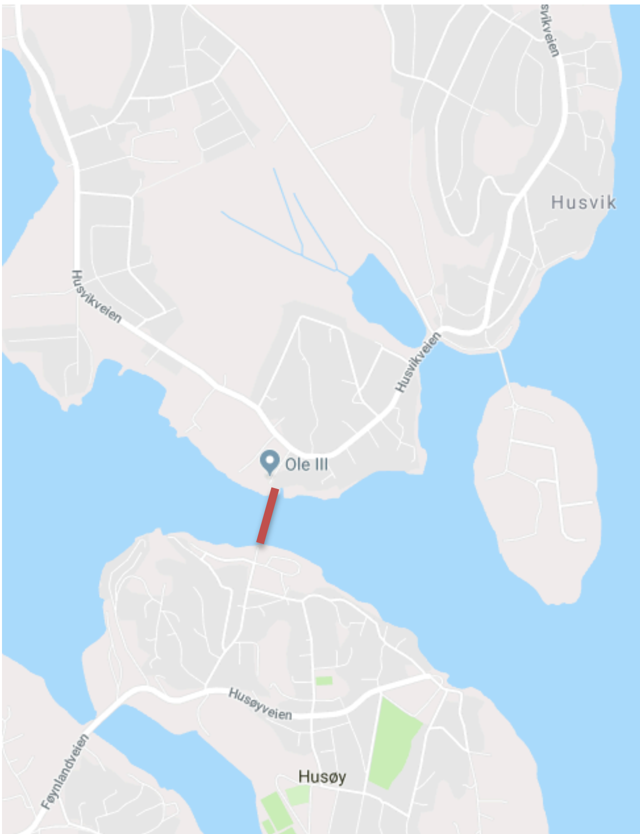

Figure 14 Ole III Vessel (Top) and its operational area (bottom) (Google, 2019)

3.3.2 Observational Report

During the regular operation of Ole III, one nautical operation bachelor student has joined the captain and filled out one observational report for each of the crossings. Different students performed data collection based on their availability. The template of the observational report can be found in Appendix A. The outcome of this study is submitted to The Applied Autonomy Systems Summit 2018 conference (USN, 2018) as a research article.

3.3.3 Video Recording

Two sets of cameras are located in the shore side to record Ole III crossings. Each camera set is consisting of two optical and thermal cameras. The cameras station and orientations are shown in Figure 15.

Figure 15 Cameras station (Top) and orientations (Bottom) (Google, 2019)

In summer due to some construction work at the Ole III jetty in Husøy side, Ole III has moved to the adjacent jetty. That is why there is a mismatch with google map and collected data. Figure 16 presents one example of video recordings.

Figure 16 Example of Video Recordings

Camera set one has a clear overview of Ole III crossing, and in this thesis recordings from this camera set is used. Thermal cameras can provide better input in weather situations with limited vision such as fog, darkness, rain (Rodin et al., 2018). Optical cameras recordings were sufficient in the current thesis since data collection was done in summer and optical cameras had a clear vision of the operational area.

3.3.4 Camera Specification

Camera with specification presented in Table 4 was used for video recordings.

Table 4 Camera Specification (Hikvision, 2019)

Camera Model:	Hikvision DS-2TD2636-10
Optical Image sensor:	1/2.8'' Progressive Scan CMOS
Optical Max. Resolution:	1920 × 1080
Thermal Image sensor:	Vanadium Oxide Uncooled Focal Plane Arrays
Thermal Max. Resolution:	384*288
Lens (Focal Length):	10mm Lens

The quality of the optical camera is set to below specification for video recordings:

Video resolution: 1920 x 1080 Pixels

Frame rate: 2 Frame Per Second (FPS)

3.3.5 Population

The total collected data from both tools are shown in Table 5.

Table 5 Description of collected data

Data Type	Duration	Start Date	End Date	Total Data Points
Observational Reports	2 Months	04 June 2018	04 August 2018	4803 Reports
Video Recordings	Two Months and 21 Days	06 June 2018	27 August 2018	82 Days

3.3.6 Secondary data

The data collection was done before the research design. Another research team collected the observation data while video recordings were performed in addition to the manual observations for future applications. Using secondary data in research facilitates the research by saving time and resources for data collection while enforces some limitations to the research (Sharp et al., 2002). The inaccuracy of observations performed by a human observer and technical issues affecting the video recordings are some of the limitations in this thesis.

3.4 Ethical Considerations

In the video data, the crossing traffic was recorded, and only the vessels and their types are visible. Due to cameras limited fidelity no attribute related to any specific vessel such as name and registration number were collected. In the observational data, the number of passengers and their characteristics (e.g., pedestrian, bicyclist and demand for special help) was collected. However, personal data about any specific individual, which could identify them were not gathered or used for data analyses and anonymity, and privacy criteria were satisfied.

In order to confirm the ethical and legal framework, the research was reported to the Norwegian Data Protection Authority (DPA), and it was approved (DPA, 2018).

3.5 Validity and Reliability

Validity is defining appropriateness of the tool used for the measurement. Validity is consisting of content (face and sampling), empirical and construct. Face validity is achieved by subjective assessment of tool utilized for the research (Frankfort-Nachmias, 2015). In the current thesis dependent variable is YOLO detection accuracy and independent variables are the number of training inputs and traffic situation. The observer onboard Ole III observational reports, YOLO and manual video observations of video recordings are used as instruments (tools) in this research. Measurement is defined as counting the numbers of vessels in the experiment setting. Due to the simple nature of measurement, face validity is considered as validity criteria.

Reliability defines to what extent an instrument can produce consistent measurements. There are different methods such as test-retest method, parallel-forms technique and split-half for performing reliability check. In parallel-forms technique, different instruments are developed for measurement and reliability is assessed comparing the results. In the split-half method, the sample is divided into subsamples, and the tool is measuring the dependent variable in subsamples. By comparing the measurement in subsamples, the reliability is achieved (Frankfort-Nachmias, 2015). In this thesis, two different methods of split-half method and parallel-forms are applied simultaneously for fulfilling reliability (see Figure 17). Form a different perspective, the reliability process can be seen as methodological triangulation across the methods (Bekhet & Zauszniewski, 2012).

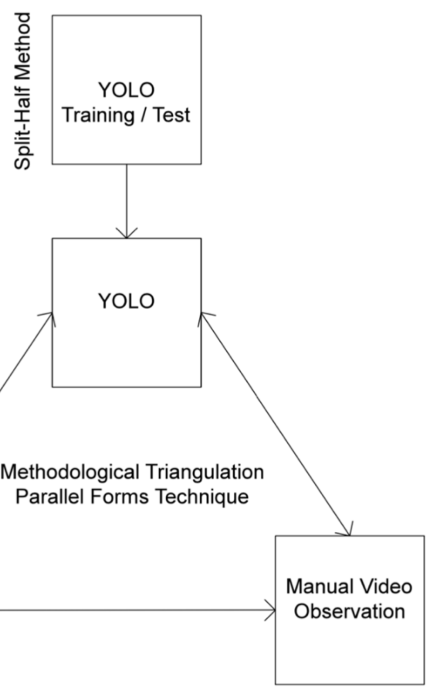

Figure 17 Reliability in Research

3.6 Data Analysis

3.6.1 Sampling

3.6.1.1 Sampling Technique and Criteria

The total amount of collected data (population) consists of 4803 observational reports and 82 days of video recordings. In order to limit the amount of data to the controllable range, sampling has to be done. Purposive sampling is used when there is a specific criterion to limit the population to sample (Frankfort-Nachmias, 2015). In this sampling method, the researcher uses its judgment for sampling which is the case in the current thesis. Depending on the navigational situation, the vessels have to give away to each other and COLREG is formulating it (IMO, 2003). Assuming the criteria "Violating of COLREG by passing traffic distinguished by the observer onboard Ole III" population was narrowed down to the sample. By implementing this criterion, the number of the data points were reduced to 183 from 4803 crossings.

3.6.1.2 Failures in Sampling

Video recordings started from 6th of June instead of 4th of June and six number of crossing videos in the mentioned time interval is not available. Due to technical issues, eight number crossings are not recorded, and that also indicates the importance of hardware reliability in autonomous operations. In one of the crossings, Ole III is standing still while observer onboard reported it as a crossing. The mentioned data points are excluded from the sample due to issues as mentioned above. The total number of 168 crossings were available from both observational data and video recordings.

3.6.1.3 Sample Distribution to Training, Test and Evaluation Datasets

The sample has been divided into three parts: training, test and evaluation datasets. Dividing data to two part of training and test is called the split-test method in the machine learning context (Goodfellow, 2017). Split-half is also a method for achieving reliability in research method area (Frankfort-Nachmias, 2015). There is no specific principle about the division sizes while the most common breakdowns are 90% - 10%, 80% - 20% and 70% - 30% for training and test respectively (Goodfellow, 2017). Computer vision algorithms are dividing the data into two separate part of training and test; however, in the current thesis research design is different due to the requirement of evaluation data. In this thesis, 60% of data is used for the preparation of YOLO training dataset. Performance of YOLO during the training phase is assessed by 20% of data which is called test dataset. The last 20% of data is used to evaluate the accuracy of YOLO by comparing with the observational data collected by the crew and by manual video observation (see Table 6).

Table 6 Training, Test and Evaluation Datasets Distribution

Data Type	Percentage	Number of Data Points (Crossings)
Training	60%	101
Test	20%	33
Evaluation	20%	34
Total	100%	168

3.6.1.4 Class Definition

In order to prepare the training and test datasets, videos were observed manually. Situations with new objects, different angles of the same object and complicated traffic situation were captured (845 images). By observing the traffic situation, crossing vessels are grouped into five classes of Ole, Motorboat, Sailboat, Rowboat and Other. The definitions and details of each group are shown in Table 7.

Table 7 Object Detection Classes (All Phases of Training, Test and Evaluation)

Class Number	Class Name	Description
0	Ole	Ole III
1	Motorboat	All the vessels with engines for propulsion including jet-skies
2	Sailboat	Sailboats sailing or using engines
3	Rowboat	Vessels with human energy as propulsion
4	Other	Utility, construction, and passenger vessels

3.6.1.5 Objects Distribution in Training and Test Datasets

The selected images were reviewed and the ones with the same content were removed from the dataset. The final dataset contains 442 number of images divided by 347 for training and 95 for the test. The object distribution and related classes in the training and test dataset are given in Table 8 and Figure 18.

Table 8 Object Distribution in Training and Test datasets

Description	Number of Images	Class Content (Number of Objects)					
		Ole	Motorboat	Sailboat	Rowboat	Other	Total
Training Dataset	347	346	1147	56	14	6	1569
Test Dataset	95	95	375	5	0	0	475

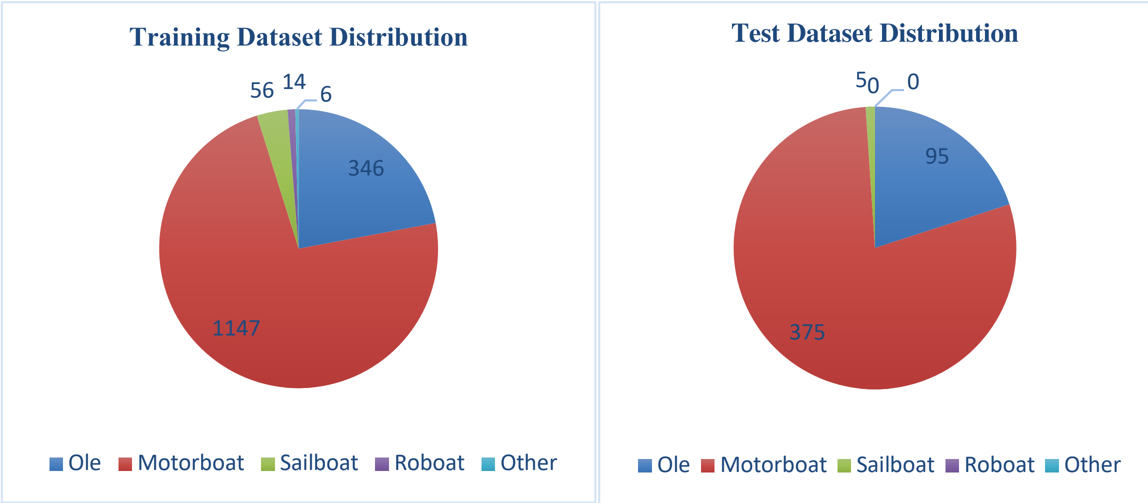

Figure 18 Training and Test Datasets Objects Distribution

3.6.1.6 Training and Test Inputs Preparation

YOLO training and test inputs need a specific format. For each of the inputs, one text file and the corresponding image has to be available. Text input has to contain the class of object, the relative location of object centroid horizontally and vertically and width and height of the object relative to overall image size for each object (see Figure 19). Each of the images was manually observed, and the availability of objects inside them was manually defined, and bounding boxes were drawn. The images were manually labeled by using the YOLO-MARK tool (AlexeyAB, 2019b) as shown in Figure 20.


```
0 0.348828 0.372917 0.053906 0.048611
1 0.695703 0.381944 0.066406 0.041667
1 0.767969 0.470139 0.103125 0.087500
1 0.706641 0.451389 0.121094 0.080556
1 0.560938 0.413194 0.042188 0.026389
2 0.489844 0.344444 0.059375 0.150000
3 0.592188 0.445139 0.060938 0.029167
4 0.807813 0.397222 0.132812 0.108333
```

Figure 19 Sample of training input (Image and Text file)

Figure 20 Example of object labeling using YOLO-Mark tool (AlexeyAB, 2019b)

3.6.2 Computer Specification

Training and utilizing of the neural networks demand high computational power. For performing YOLO, a computer with specification presented in Table 9 was allocated for this task.

Table 9 Computer Specification

Operating System	Ubuntu 16.04 LTS
Memory	31,4 Gib
Processor	Intel® Core™ i7-8700K CPU @ 3.70GHz × 12
Graphics	GeForce GTX 1080/PCIe/SSE2
OS Type	64-bit
Disk	330,3 GB

4 Results

The YOLO algorithm was trained based on the crossing traffic and evaluated for future applications. During the training phase, YOLO accuracy is calculated on test dataset in order to have control over the performance. In the next phase, evaluation dataset is analyzed using three methods of observer onboard the Ole III, YOLO object detection system and manual video observation. By comparing the outputs, different methods were evaluated. Detections are classified into five groups (see Table 7), as described in the data analysis section.

4.1 Accuracy in training phase of the model

The current research focusses on the suitability of YOLO v3 algorithm for the object detection task in the maritime operations. That is why the capacity of the model in this context was not modified.

In the current research, the training phase is divided into two phases of training and test. The training data set is used for updating the parameters in the neural network and lowering the cost function while at the same time test dataset is used to evaluate the accuracy of the detections. On each epoch, the training data is given to the network, and the cost is calculated. In the backward phase, the parameters are updated, and at the end of each loop, accuracy is calculated. The cost function values and accuracies are drawn in Figure 21. By selecting the point with a lower number of epochs, the chance of underfitting is increased while by performing more epochs, the model is more likely to overfit. Selection of appropriate point is a critical decision in research.

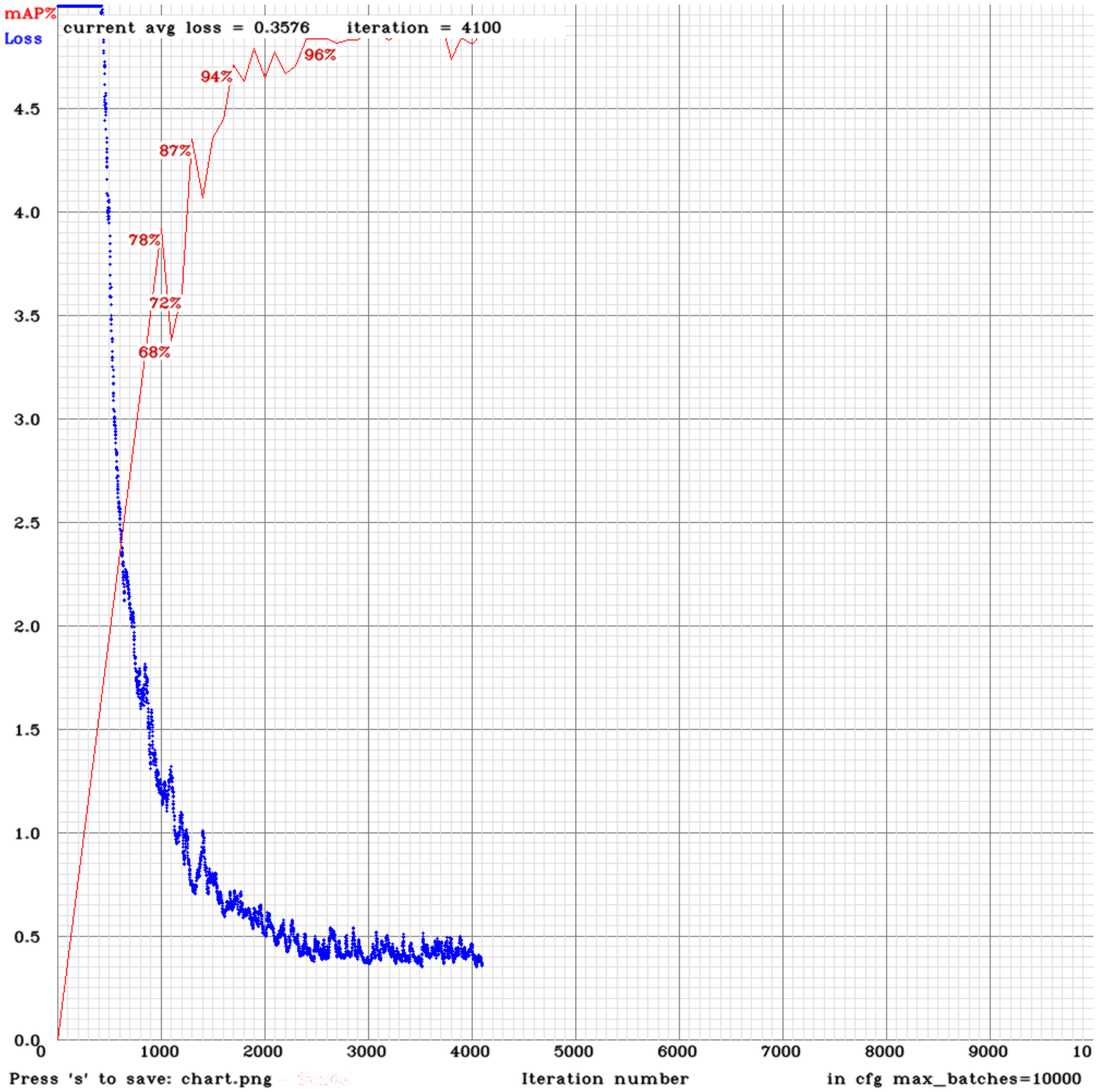

Figure 21 Overall loss vs. mean Average Precision (mAP) in training phase

The model weights are stored in every 1000 epochs. As can be seen from the curve, until 2000 epochs, the rate of increase in accuracy and drop in loss is substantial. After the 2000 epochs, the change rate is reduced considerably, and the curves are flattening. Epoch 2000 is the elbow point in the curve and considered as a good fit for the model. In the epoch 2000, model has the accuracy of 94% on the test dataset.

4.2 Evaluation Dataset

One of the main criteria in the object detection systems is that the system should not see the evaluation data. That is why the 34 crossings were not used in any previous stages of the system development.

4.3 Total Number of Detections

4.3.1 Observer Onboard Ole III

The observer onboard was asked to report the intervening traffic which will affect the navigation of Ole III. In the data collection phase for each of the crossings, one observational report was filled. For the evaluation crossings, corresponding reports were collected and processed in order to understand the nature of the traffic. In the observational report, there is no recording for the Ole III as the observer is onboard the vessel and detection of Ole III is not applicable. The summary of observer onboard Ole III detections is presented in Table 10 and Figure 22.

4.3.2 You Only Look Once (YOLO)

For assessing YOLO performance, evaluation videos were given to the algorithm, and the detections were reported. The detection criteria in YOLO is to report any moving vessel which is entering or exiting the camera field of view. Detection has to be stable, and momentarily detections are not considered appropriate. The camera has an overview of two jetties where some vessels are parked there seldom. The system detects the stationary vessels while they have not been reported due to the detection criteria. The summary of YOLO detection is presented in Table 10 and Figure 22.

4.3.2.1 Misclassification and Misdetection Errors

Two types of errors are possible on YOLO detections. First is the situation that YOLO is unable to detect an object which is considered as misdetection error. Second is YOLO is detecting an object while classifies it as a wrong class which is misclassification error.

4.3.3 Manual Video Observation

Observational reports and video recordings were compared together, and some differences realized. In order to develop a baseline (ground truth) for the evaluation phase, the video recordings were manually reviewed by the single observer (the researcher). The detection criterion was the same as being used in YOLO detections. Moving vessels in cameras field of view which are entering and exiting the frame were detected. The summary of manual video observation is presented in Table 10 and Figure 22.

Table 10 Total Number of Detections for Three Methods per Class on Evaluation Dataset

Method	Ole Class	Motorboat Class	Sailboat Class	Rowboat Class	Other Class	Total
Observer Onboard Ole III	34	65	6	1	1	107
YOLO	34	94	7	1	1	137
Manual Video Observation	34	91	7	4	2	138

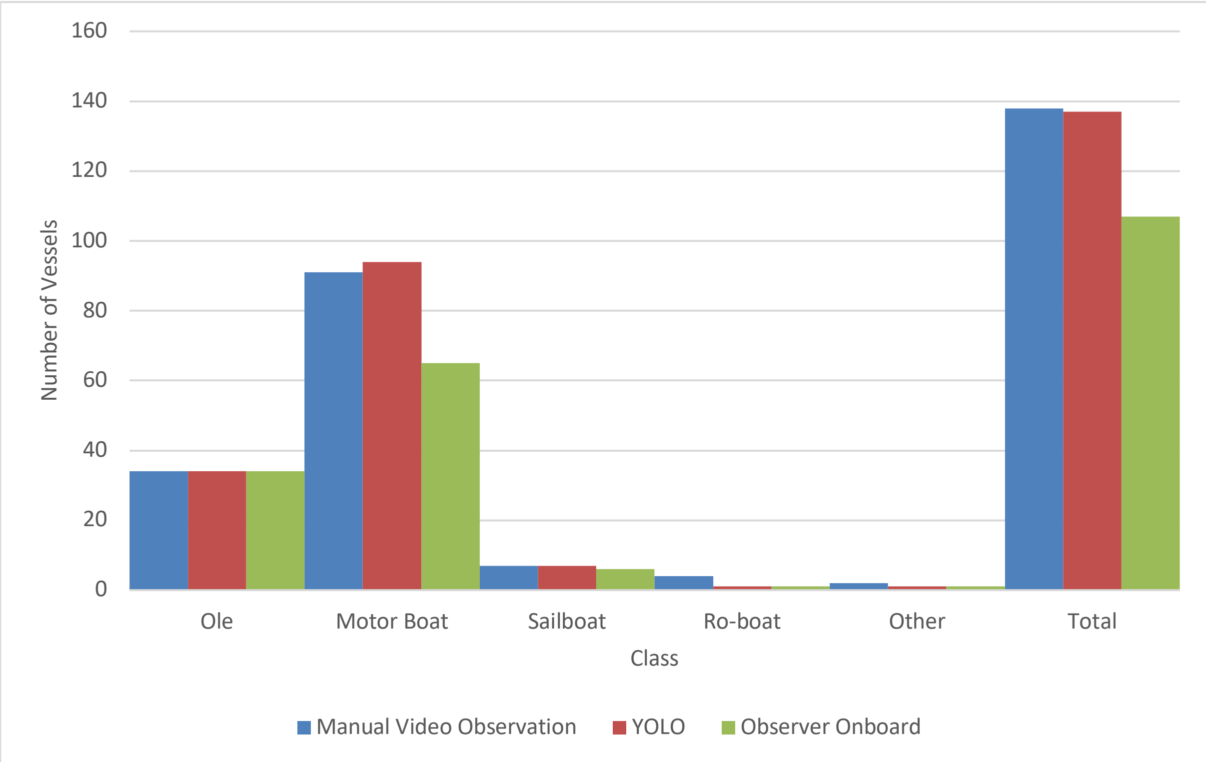

Figure 22 Overview of Total Vessels Detections per Class from Three Methods on Evaluation Dataset

4.4 Comparison

The overview of detections on evaluation dataset (34 Crossings) for three methods of Observer Onboard Ole III, YOLO, and Manual Video Observation were presented in the above section. In this section performance of each method for specific class detection is discussed, and at the end, the total number of detections per crossing is presented.

4.4.1 Ole Class

Evaluation of Ole III from observer onboard perspective is not applicable as the observer is sailing with the vessel. The camera has a clear overview of Ole III operational area, and the vessel is visible in all the recordings. From manual video observation perspective, this evaluation is not applicable. YOLO detected Ole III in all phases of operation and detection accuracy is 100% in this case.

4.4.2 Motorboat Class

The number of motorboats detected per crossing for each method is introduced in Figure 23. The vertical axis is showing the number of vessels passing, and the horizontal axis is the specific crossing number.

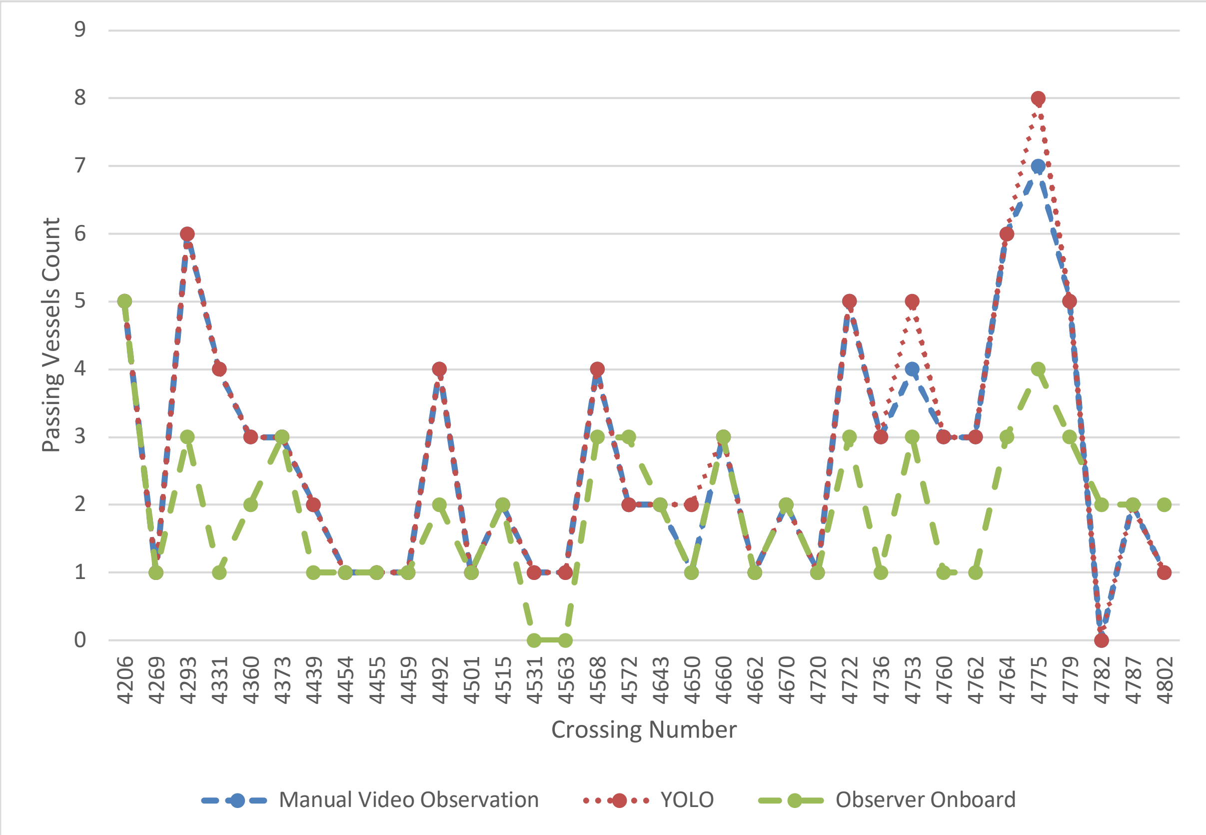

Figure 23 Motorboat Class detection per crossing on Evaluation Dataset

It can be seen from the figure that manual video observation and YOLO detections are matching each other to a great extent except crossing numbers 4753 and 4775. In both of the crossings, one of the vessels passing is rowboat which YOLO detects it as motorboat (see Figure 24) which is misclassification error. The same error will also reflect in rowboat class detection.

There is a considerable difference between observer onboard evaluation and YOLO and manual video observations.

Figure 24 Misclassification errors on crossings 4753 (Bottom) and 4775 (Top)

4.4.3 Sailboat Class

The number of sailboats passing per crossing on evaluation dataset is presented in Figure 25 for each method.

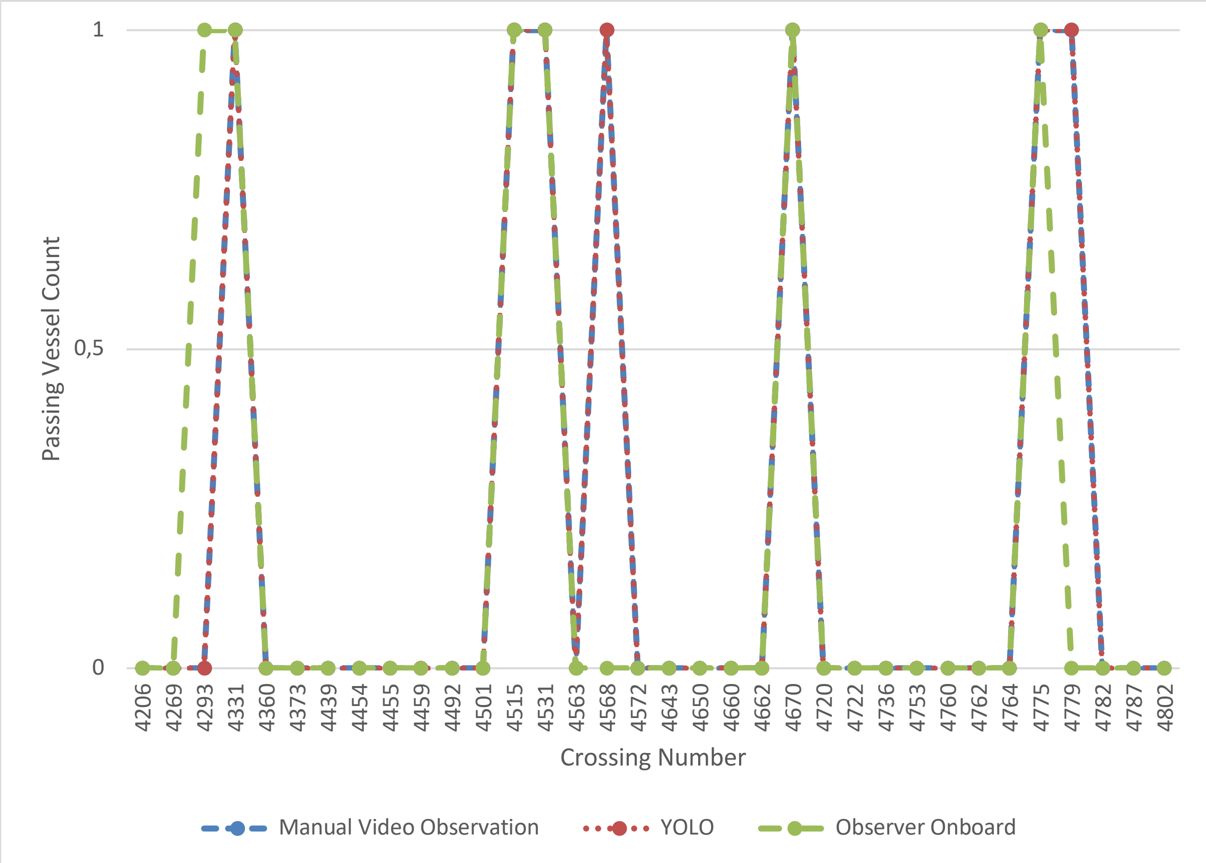

Figure 25 Sailboat Class detection per crossing on Evaluation Dataset

It can be seen from the above graph that the detection of three different systems matches each other to a great extent except three crossings. In crossing 4293, observer onboard detects one sailboat passing while YOLO and manual video observation were unable to detect it. In crossings 4568 and 4779, one sailboat was detected in manual video observation and YOLO detection while it was not reported by observer onboard the vessel.

4.4.4 Rowboat Class

The detection of rowboat class per crossing for three methods is shown in Figure 26.

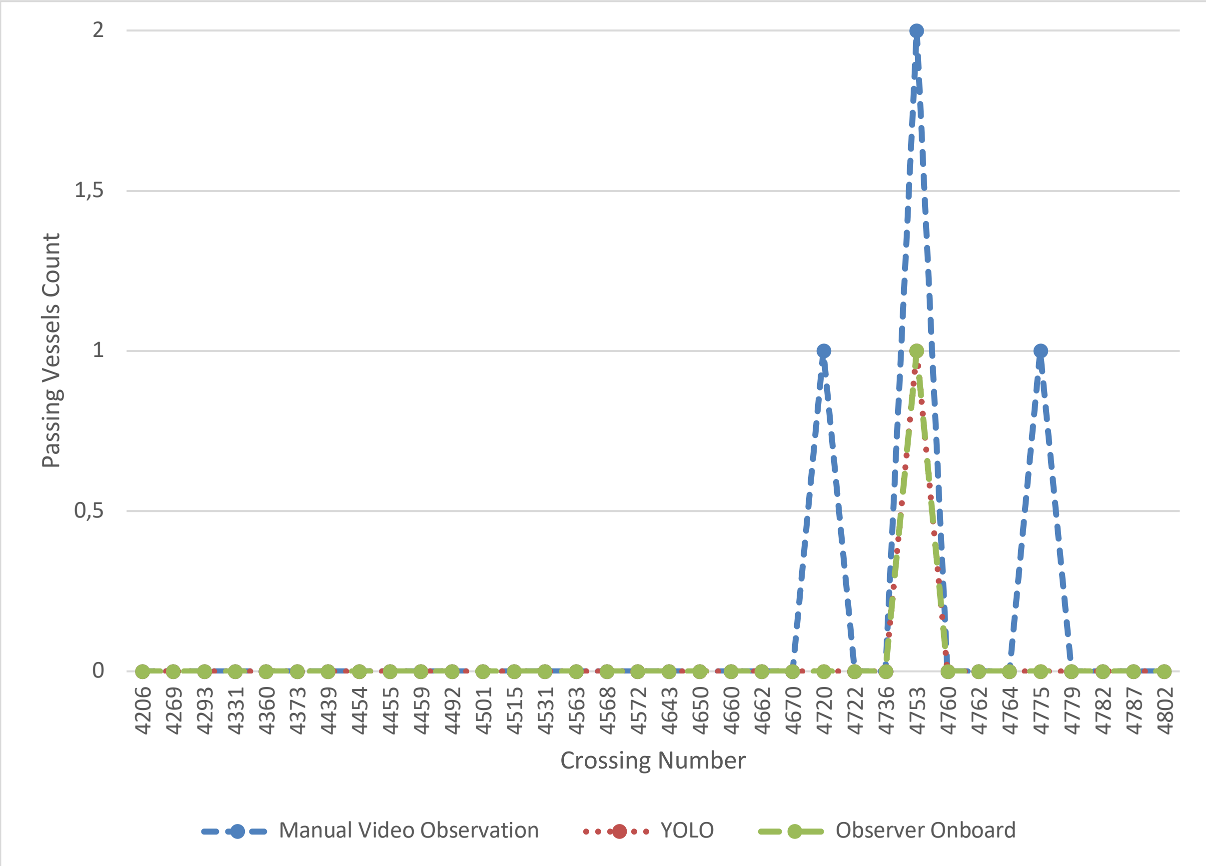

Figure 26 Rowboat Class detection per crossing on Evaluation Dataset

It can be seen in most of the crossings there is no rowboat in this area. In the crossings 4720, 4753 and 4775 rowboats are passing. In crossing 4720, one rowboat with four paddlers was detected by manual video observation while YOLO and observer onboard did not detect this vessel and it is a misdetection error in YOLO (see Figure 27). During the preparation of YOLO inputs in training and test datasets, such vessels were not seen. In crossing 4753 two rowboats has been detected by the manual video observations while both YOLO and observer onboard only reported one of them (see Figure 24). The YOLO detect the second rowboat as motorboat which is misclassification error. For the crossing 4775, the manual video observation detects one rowboat while YOLO and observe onboard did not detect any rowboat (see Figure 24). YOLO detects mentioned rowboat as motorboat which is a misclassification error.

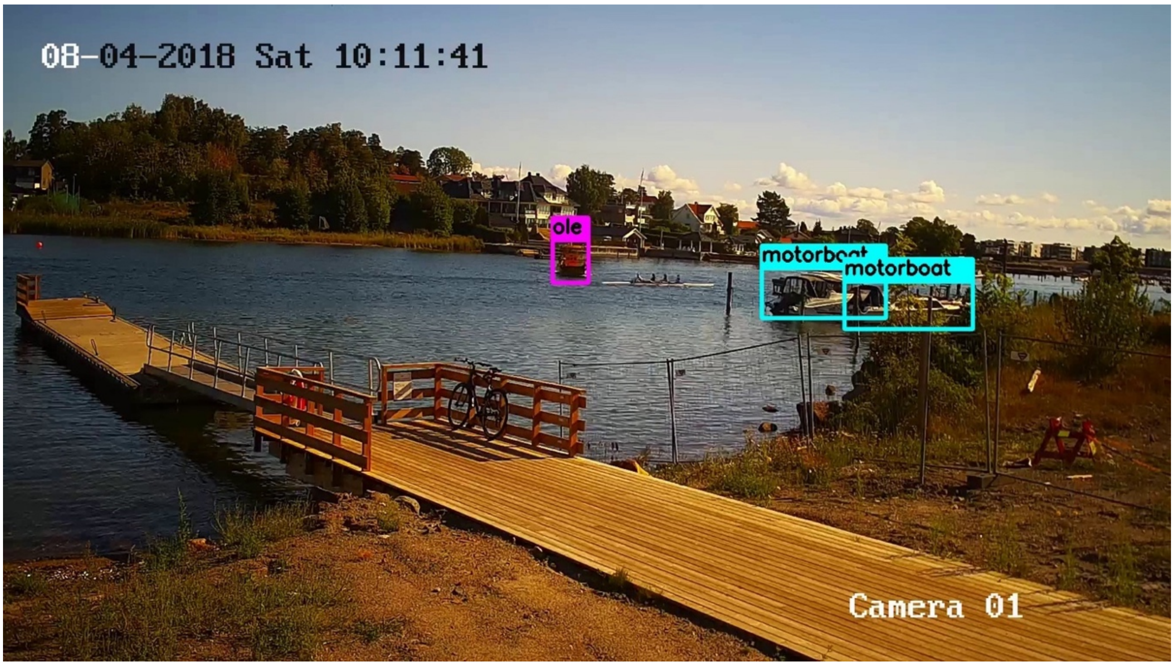

Figure 27 Misdetection of Rowboat class in crossing 4720

4.4.5 Other Class

Other class detections per crossings for three methods are shown in Figure 28.

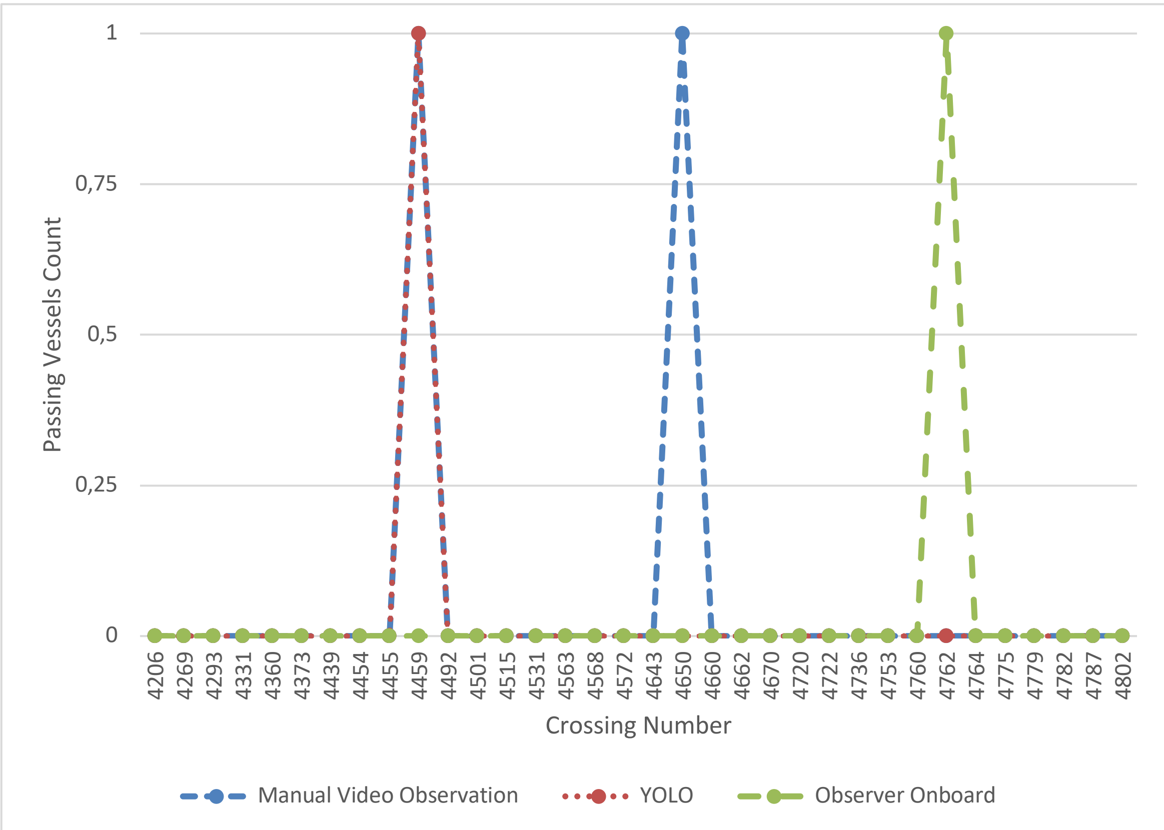

Figure 28 Other Class detection per crossing on Evaluation Dataset

There is low traffic for this class type in the research location. In crossing 4459, one passenger vessel is crossing, both manual video observation and YOLO have detected this vessel while observer onboard has not reported it. In crossing 4650 (see Figure 29), one rescue vessel is crossing, and manual video observation detects it as other class while YOLO detects it as a motorboat, and that is why it is considered as misclassification error. On crossing number 4762, observer onboard Ole III detects one of the crossing vessels as other class while YOLO and manual video observation did not detect it.

Figure 29 Other class misclassification in 4650 crossing

4.4.6 Total Number of Detections

The total number of detections per crossing for each method is presented in Figure 30.

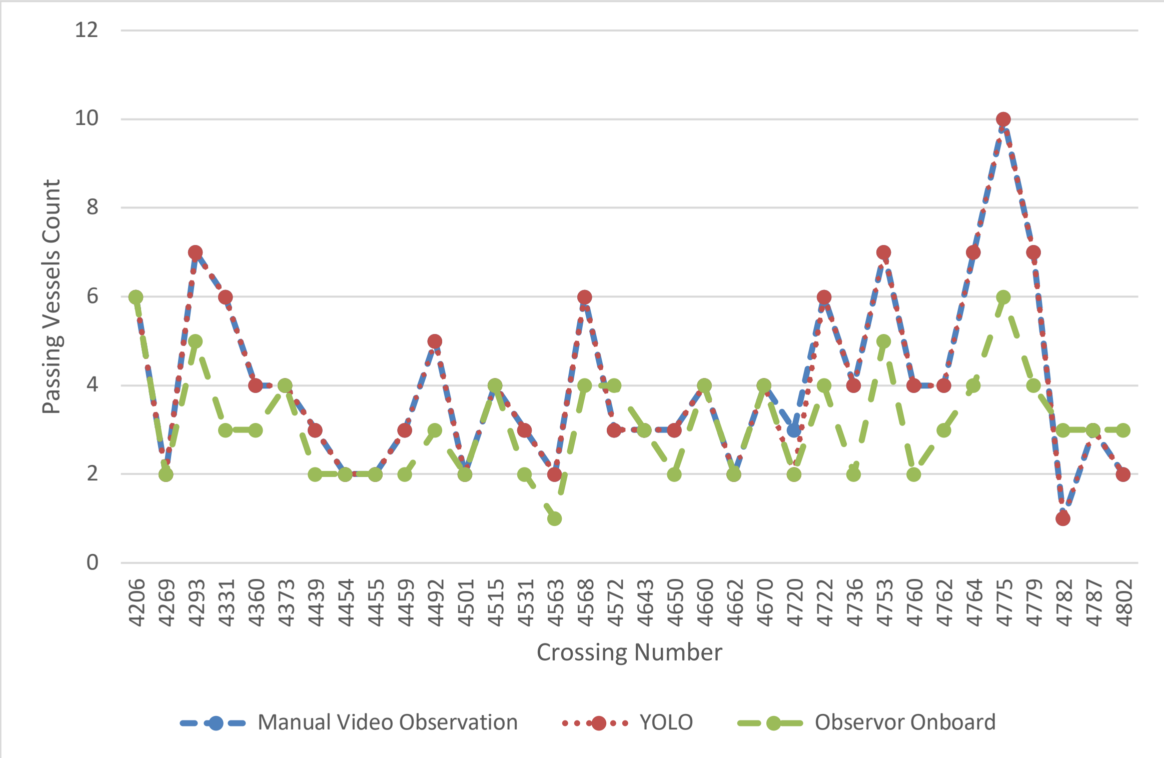

Figure 30 Total Number detections per crossing on Evaluation Dataset

As it is shown in the graph, YOLO and manual video observation detections fit well except the crossing 4720 which YOLO is unable to detect one rowboat (misdetection). The misclassifications of YOLO cannot be seen in this curve. Misclassification error happens when the vessels are assigned to the wrong class, and it does not reflect in the total number of detections. The observer onboard in most of the crossings under evaluate the number of the vessels passing.

5 Discussion

One of the main goals of developing this research was to assess the suitability of YOLO computer vision algorithms in the maritime context. Traffic nature of evaluation dataset were analyzed using three different methods of observer onboard reporting, YOLO and manual video observation. By comparing the YOLO detections with observational methods, the accuracy of YOLO is calculated.

5.1 YOLO Accuracy

5.1.1 Baseline

The main task of the observer onboard Ole III was to assess the violations of COLREG as well as studying the decisions made by Ole III captain in those situations. Therefore, the number of passing traffic was only recorded in the case of affecting the vessel navigation. Vessels in the distant locations might appear in the cameras field of view while they do not have an effect on Ole III navigation. The above-mentioned points might affect the consistency of the data collected by the observer onboard Ole III and video recordings.

Video recordings were reviewed by a human observer (the researcher) in order to evaluate the number of passing vessels. Comparison of the results between manual video observation and the observer onboard showed that there are some differences on evaluation between these two methods. YOLO and manual video observation are having same setting for evaluation and using manual video observer evaluations as baseline is more rational. Manual video observer might also fail to perform the detection task correctly due to the human errors (Woods, Dekker, Cook, Johannesen, & Sarter, 2017). In order to minimize the manual video observation errors, the detections were re-evaluated multiple times with different settings. In addition, another researcher performed the manual video observation task and achieved the same results.

5.1.1.1 Statistical Testing

Statistical tests are comparing an attribute of the distributions such as mean, median or rank (Navidi, 2006). However, in the thesis settings each of the crossing is a unique distribution and comparison has to be performed within the crossing.

Reviewing the results in the previous section indicated that there is a substantial difference between the manual video observation and the observer onboard detections. The manual video observation is a more precise source in comparison to observer onboard reporting and it will therefore be used as the baseline (ground truth) for YOLO accuracy calculation.

5.1.2 Criteria

Two error types of misclassification and misdetection are possible for YOLO accuracy calculation, and both of them are considered as a fault in the accuracy computation. YOLO accuracy is estimated by dividing the number of YOLO correct detections by the number of objects detected by manual video observations.

The details about the availability of different classes in training and test datasets and YOLO detection accuracy on evaluation dataset can be found below.

5.1.3 Ole Class

For the detection of the Ole class vessel, 346 number of input objects were used in the training dataset while 95 number of inputs were used in the test dataset. In the evaluation phase, the model reached an accuracy of 100% for this class detection. Since Ole III is present in all the inputs of the training phase and it has a unique shape, this class has a high detection accuracy.

5.1.4 Motorboat Class

For the detection of the Motorboat class vessels, 1147 and 375 number of inputs were used in the training and test datasets, respectively. In the evaluation phase, the model achieved an accuracy of 96.70% for this class detection. A Considerable amount of inputs is available for the Motorboat class on the training dataset, which explains the high detection accuracy for this class.

5.1.5 Sailboat Class

For the detection of the Sailboat class vessel, 56 number of inputs were used in the training dataset while 5 number of inputs were used in the test dataset. In the evaluation phase, the accuracy of 100.00% was obtained for this class detection.

Computer vision algorithms perform the detection based on the distinguished features. Some object has more specific characteristics compared to others which makes them easier to detect by the system (Kapur, 2017). Although the Sailboat class has a small portion in the training dataset, it leads to a high accuracy for the model. This can be explained by the unique shape of this class of vessels (having a mast).

5.1.6 Rowboat Class

For the detection of the Rowboat class vessels, 14 number of inputs were in the training dataset and no inputs were used in the test dataset. In the evaluation phase, the model achieved an accuracy of 25% for this class detection. The limited number of Rowboat class vessels in the training dataset rooted in the traffic nature of the area.

5.1.7 Other Class

For the detection of the Other class vessels, 6 number of objects were used in the training dataset whereas no inputs were utilized in the test dataset. In the evaluation phase, the model attained an accuracy of 50% for this class detection. Availability of the Other class data in the training phase is low, which can be attributed to the nature of the traffic in the area.

5.1.8 Total Number of Detections

The total number of 1596 inputs were used in the training phase while 475 number of inputs were used in the test phase. YOLO achieved the overall accuracy of 94.93% in the evaluation dataset. The summary of the training and test dataset distributions per class and YOLO accuracy is presented in Table 11. Figure 31 compares the contribution comparison of different classes to the training dataset as well as the detection accuracy for each class.

Table 11 Summary of Training and Test Dataset Distributions and YOLO Accuracy

Class	No. of objects (Training dataset)	Percentage contribution (Training dataset)	No. of objects (Test dataset)	Percentage contribution (Test dataset)	Detection Accuracy (Evaluation dataset)
Ole	346	22.05%	95	20.00%	100.00%
Motorboat	1147	73.10%	375	78.95%	96.70%
Sailboat	56	3.57%	5	1.05%	100.00%
Rowboat	14	0.89%	0	0.00%	25.00%
Other	6	0.38%	0	0.00%	50.00%
Total	1569	100.00%	475	100.00%	94.93%

Figure 31 reveals that YOLO had harder times detecting the classes, which did not have sufficient contribution to the training dataset whereas providing a sufficiently large dataset resulted in higher accuracy in detection.

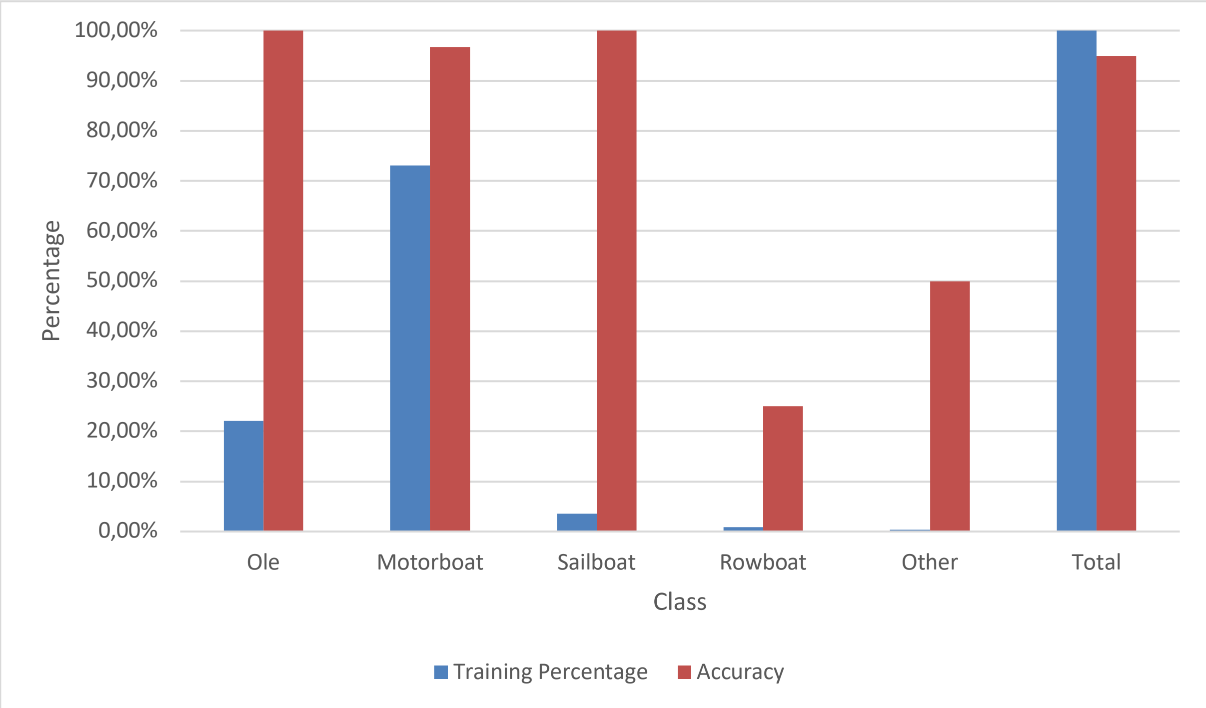

Figure 31 Comparison between the Percentage contribution to the training dataset and YOLO Accuracy for different class detections

The distribution of the entire dataset of all the classes implies that there are fewer numbers of the sailboats, rowboats and the other class vessels in comparison to the motorboats and that is dictated by the nature of the traffic in this area. In order to increase the accuracy of the detection, it is possible to increase the amount of the training data for the classes with lower contribution while it will affect the research design.

5.2 Observer Onboard Performance

In order to determine the performance of the observer onboard, the detections are compared with the manual video observations, which was chosen as the baseline earlier. Figure 32 illustrates the relative deviation between the total number of detections performed by the observer onboard and the manual video observation. Calculation basis used for estimation is the same as used for YOLO accuracy calculation. The observer onboard main objective was to assess violations of COLREG and detecting the number of vessels passing was defined as a secondary objective. In other words, our knowledge about the assumptions made by the observers onboard is not adequate to make a solid argument about their detection. Thus, Figure 32 does not indicate the accuracy of the observer onboard detection and is merely presented for visualization.

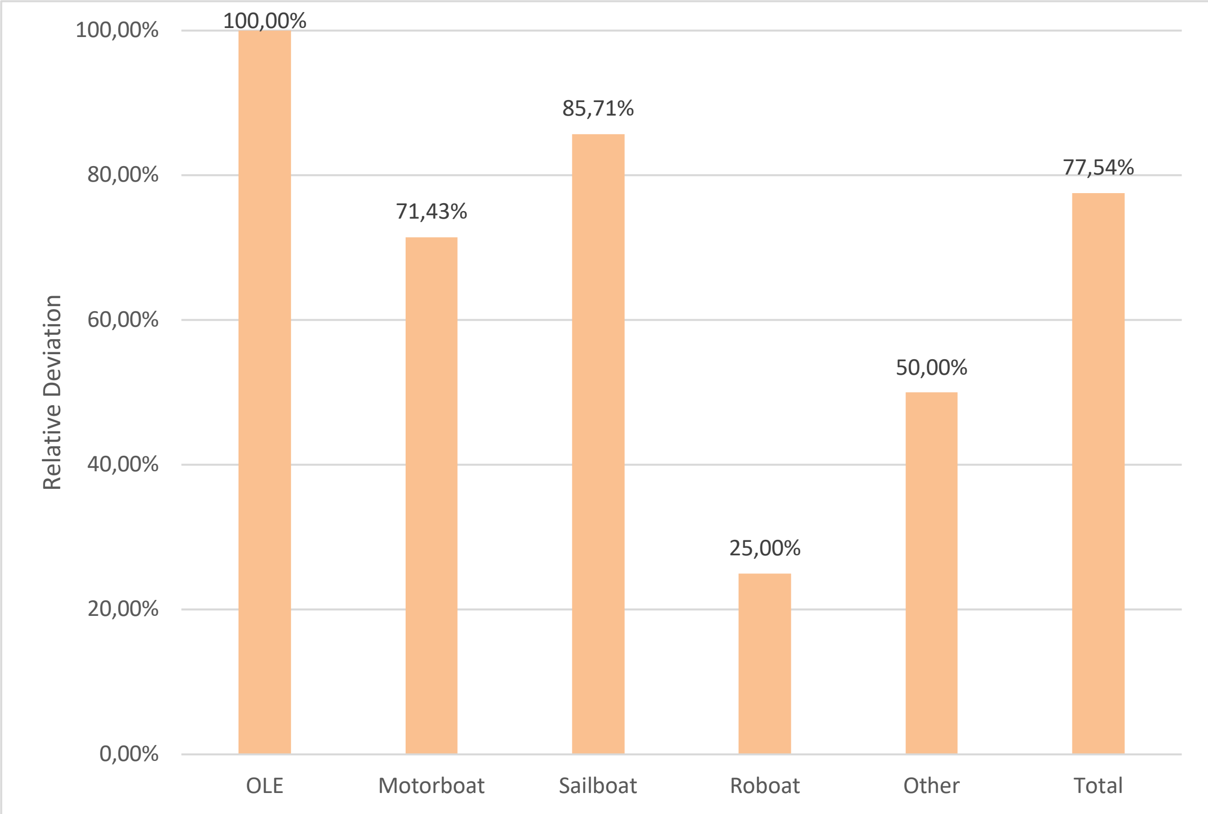

Figure 32 Observer Onboard detection comparison with Manual Video Observation

5.3 Validity and Reliability

Validity is defined as the appropriateness of the tool used for the measurement (Frankfort-Nachmias, 2015). In this research accuracy of YOLO is considered as the dependent variable while the number of training inputs and the traffic situation are independent variables. In order to assess the validity of the research, face validity method is chosen. In this study, measurement consists in performing a detection as well as classification the passing traffic using three tools of the observer onboard, YOLO and the manual video observer. The detection task was simply done by the counting the number of the vessels passing in the research area and classifying them into groups. YOLO executes the detection task based on the technical steps presented in the background chapter. Counting the vessels and classifying them are possible using the tools discussed and validity can be assessed within the face validity context.

Reliability refers to what extend an instrument can produce consistent measurements (Frankfort-Nachmias, 2015). This study has showed that there are some inconsistencies associated with the observer onboard evaluation. The observer onboard reporting are secondary data and the investigation of the source of inconsistencies is beyond the scope of the current research. Given these issues, the observer onboard evaluation cannot be used in the evaluation stage as baseline. Therefore, the manual video observation was selected as the ground truth for YOLO evaluation due to having same settings. It should be noted that the manual video observer was also prone to the human related errors (Woods et al., 2017), however, controlling the experiment setting and multiple reviews of videos alleviated this type of limitation. In order to fulfil reliability requirement, two different methods of split-half and parallel-forms techniques are utilized simultaneously. Split-half method was used in the training phase of the neural network where data was divided into two parts of the training and test datasets. In epoch 2000, network achieved the overall accuracy of 94.00% on the test dataset (see Figure 21). In parallel-forms technique the accuracy of YOLO was compared with the manual video observation evaluation. Analyzing the evaluation dataset indicated that the network has the overall accuracy of 94.93% (see Table 11). As a result, the detection accuracy is obtained within the same range using both approaches.

5.4 Comparison with Benchmarks

To assess YOLO accuracy, it has to be compared with benchmarks. However, computer science benchmarks such as PASCAL VOC (PASCAL, 2019) and COCO (COCO, 2019) datasets are not suitable for comparison. These datasets are produced for general-purpose detection systems, and no specific dataset is available for maritime applications. Furthermore, these datasets provide detection and localization accuracy combined as a single value while the current research is only concerned with detection. As a result, such comparison between the accuracies is not legitimate.

As reported in the autonomy literature, Rodin et al. (2018) used a CNN algorithm for object detection during a maritime search and rescue operation using the data collected by an unmanned aerial vehicle and achieved 92.5% detection accuracy. In addition, Yang-Lang et al. (2019) implemented YOLO v2 for target detection in SAR images for maritime traffic monitoring and achieved 90% detection accuracy. The current study used YOLO v3 for the detection task and attained an accuracy within the same range using a somewhat similar setting.

5.5 Misclassification and Misdetection Errors

The YOLO detection errors are classified into Misclassification and Misdetection errors. Misclassification happens when YOLO assigns a vessel to a wrong class while Misdetection is associated with the situations when YOLO is unable to detect a vessel. In the accuracy calculation both error sources were considered as system faults. Assuming YOLO is providing the sensory data to the collision avoidance system on an autonomous vessel two different consequences levels are anticipated. In the case of misdetection more severe consequences are possible comparing to the misclassification. As an example, misdetection of a rowboat with four paddlers has more destructive ramification than misclassification of a rescue vessel as a motorboat.

5.5.1 Safety

Safety is one of the main concerns for the autonomous vessels operations (Thieme et al., 2018). YOLO algorithm has achieved 94% accuracy on the test phase and 95% accuracy on the evaluation phase. These numbers are perceived to be within the same range as the benchmarks while the errors must also be examined from the safety perspective. The errors took place in those groups, which are prone to more severe consequences in the case of accidents. This raises the safety concerns for the implementation of this system. Further research is required to study the probabilities and consequences of accidents from the safety point of view.

5.6 Traffic Situation

The traffic situation is considered as the independent variable to evaluate the accuracy of YOLO, which is the dependent variable. Traffic is an inherently stochastic phenomenon, which is difficult to be described by a specific pattern (Kim et al., 2018). In this case, the traffic situation of the area under experiment mostly consists of leisure vessels. Machine learning algorithms predicts the future incidents based on the historical data (Goodfellow, 2017). However, any new or abnormal happening, which are not included in the training phase, lead to failure in the performance or false results. For instance, a construction vessel (dredger) was observed in the same area but this observation was not a part of the targeted sample. In addition, not even one swimmer was reported with the experiment setting despite the fact that swimming is a favorite activity in this area in summer. Consequently, implementation of YOLO for collision avoidance system on Ole III may result in hazardous accidents such as the collision with a dredger or a swimmer with serious ramifications for passengers, swimmers and the vessels.

The population under study is highly unpredictable due to the stochastic nature of the traffic, a purposive sampling was applied to narrow it down to a sample concerning with our research objectives. It is worth mentioning that purposive sampling is a nonprobability method where all the participants do not have the same probability of being selected (Frankfort-Nachmias, 2015), but it does not change the stochastic nature of the participants inside the targeted sample. This study approximated some patterns in the traffic of the area under study as shown in Figure 22, but further research is required to fully model the traffic flow.

5.7 Research Questions

5.7.1 Research Question 1

Does the developed computer algorithm (YOLO) comply with the observation reports gathered by the crew during watchkeeping?

This thesis does not have any control over the observational reports since they were used as secondary data sources. Several differences were observed between the observational report and the video recordings, which adversely affects the consistency of the evaluations between these two methods. The current research was unable to explain the sources of observed inconsistencies between mentioned methods. In order to achieve this goal, an improved set of observations, recollected in a more systematic way, is required.

5.7.2 Research Question 2

To what extent is the machine learning algorithm (YOLO) accurate for application in autonomous vessels operations?

For the setting under study, YOLO achieved a detection accuracy within the same range as the benchmarks, which makes it a promising alternative for practical applications. However, in order to realize the application of such solutions, it is critical to take safety considerations into account. The most crucial step to take to achieve a high accuracy is the training data preparation step, which must be appropriately done. In this thesis, the selected training dataset was exclusively prepared for the setting under experiment. In a similar manner, the test and evaluation datasets were also prepared. Furthermore, it should be noted that although many studies have been conducted in the computer vision area, most of the datasets available are of general nature and maritime datasets have been rarely investigated. This thesis addresses this limitation by providing a new maritime-related dataset.

5.8 Application in autonomous operations

Autonomous vessels understand the surrounding environment by deployment of various sensors. Sensory inputs allow for the best possible decision-making relating to the operation of the vessels. Optical cameras are one of these sensors, which provide inputs that can be processed by computer vision algorithms in order to precisely accomplish critical tasks such as collision avoidance (Rolls-Royce, 2016).

This thesis is an attempt to investigate the appropriateness of YOLO for processing the camera sensory input besides performing the detection task. YOLO achieved an accuracy within the same range as the benchmarks, which proves its potential for further developments. In order to apply the YOLO to practical applications, the accuracy of YOLO can be improved by providing more training objects and datasets and implementing it on other settings.

5.9 Hardware Reliability

There is a broad range of concerns about the optimum operation of the autonomous vessels such as the rules and regulations, insurance, technological reliability, maintenance, operators competency, cybersecurity and safety (Komianos, 2018). Checking the evaluation dataset showed that video recordings of eight crossings were missing. The issue might be because of the failure in the system to record these particular crossings or the disconnection of the communication system to the server. This indicates the importance of the hardware reliability in the autonomous operations. The system design and its reliability were not the focus area of the research and cannot be discussed further.

5.10 Limitations

5.10.1 Secondary Data

Utilizing secondary data facilitates the research process while it imposes some limitations (Sharp et al., 2002). Amongst the crossings involved in the targeted sample, six crossings were eliminated from the experiment because in this case the video recording started two days later than the observer on board. Camera settings were fixed to a specific level, which limited the adjustments of parameters for YOLO algorithm.

5.10.2 Generalization

This thesis is only concerned with Ole III and its specific voyages. The single case nature makes the study to lose its generality (Frankfort-Nachmias, 2015). To overcome this limitation, it is possible to apply the developed YOLO algorithm to different vessels, other routes, and locations.

5.10.3 Observer Onboard Evaluation

Analyzing the evaluation dataset with observer onboard reporting indicated that there are some inconsistencies on observer onboard assessments. In some of the crossings, the observer onboard was unable to identify the COLREG violations. It should be emphasized that the main objective of the observations was to assess COLREG violations and the Ole III captain decisions, and counting the number of vessels was defined as the secondary objective in the case of affecting the navigation of Ole III. Therefore, the detection criteria used for the observational reports differ from the criteria for the detection of video recordings.

The observer onboard reported a crossing at a specific date and time whereas according to the manual video observation, the vessel was standing still, and no crossing happened at that specific moment.

There is a difference between the field of view of the observer onboard and cameras. This might influence the traffic estimation for the video recordings and observers reporting. Moreover, human observers onboard might accurately fail to assess the situation due to human errors (Woods et al., 2017).

All of these concerns question the consistency of observations performed by the observer onboard and manual video observations. Observational reports are secondary source of data and analyzing the inconsistency sources are beyond the scope of the current thesis.

5.10.4 Camera Resolution and Video Speed (FPS)

Due to the limited resources on storage and communication bandwidth, the camera recordings were set at 2 Frame Per Seconds (FPS) and resolution of 1920 × 1080 pixels. Fixed FPS limits the algorithm speed evaluation in this research. Redmon and Farhadi (2018) argues that, computer vision algorithms have lower accuracy when detecting small objects. It might be another reason for low accuracy in Rowboat Class detection. Fixed resolution of videos is another limitation of this research, which does not allow us to discuss this concern further. However, both FPS and resolution effects are out of the scope of this research.

5.10.5 Computational Power

Utilizing neural network systems are resource intensive tasks which require high computational power (Goodfellow, 2017). The training phase of the current research faced some difficulties which, were caused by the lack of computational power. Using small batch size in the training phase increased the fluctuation in neurons weight update and produced convergence issues.

6 Conclusion

6.1 Concluding Remarks

The idea of this thesis was initiated from a research project, which studies the feasibility of replacing Ole III by an autonomous vessel in the future. The current thesis evaluates the suitability of You Only Look Once (YOLO) version 3 for application in autonomous vessels operations such as collision avoidance system. In summer 2018, the crossings of Ole III were recorded by both individuals onboard and video cameras, resulting in the total number of 4803 observational reports and 82 days of video. Considering "the violations of COLREG by passing traffic assessed by observer onboard" as criteria for purposive sampling, the number of crossings were reduced to 168 crossings. Therefore, the sample was divided into three datasets of training, test, and evaluation, which include 60%, 20% and 20% of the sample, respectively. YOLO accuracy was evaluated in two phases of the test and evaluation. Two methods of manual video observation and YOLO were applied for the detection and classification of vessels through video recordings. Moreover, the observational report filled by the observer onboard was also used for comparison.

Several differences were observed between the observational report and the video recordings, which adversely affects the consistency of the evaluations between these two methods. Current research is unable to identify the sources of these inconsistencies between the methods due to utilizing secondary data. By estimating YOLO accuracy in the training phase, YOLO achieved an accuracy of 94% on the test dataset. YOLO achieved an accuracy of 95% in the evaluation phase, which is close to the detection accuracy in the training phase. Detection accuracy of YOLO was lower in the classes with lower percentage contribution in the training phase while classes with a higher number of objects in the training phase achieved higher accuracy. The accuracy obtained by the developed YOLO algorithm is within the same ranges as the benchmarks with similar experimental settings. To sum up, this work has shown that a well-developed computer vision algorithm is a promising alternative for practical applications in autonomous vessels operation such as collision avoidance. However, further research and development needed to address the safety concerns about the errors in detections. This thesis is a single case study. Therefore, in order to increase the generalization, the method has to be applied to different experimental settings. Consequently, it will be more applicable to a broader range of practical problems.

6.2 Recommendations for Future Research

In order to improve the detection accuracy, it is recommended to increase the number of training inputs. To achieve this, more crossings of Ole III has to be studied, and more inputs shall be prepared for the detection system. This case is more critical for the classes with lower training data.

The developed YOLO algorithm is only trained on Ole III and its specific voyages. It is crucial to test and retrain the current network with other settings, locations, routes, and vessels to improve the generalization of the research.

Cameras used in the current thesis experiment design were located in the fixed positions on the shore side. Stationary cameras onshore have a limited field of view. In order to gain a better understanding of the traffic situation in the area under experiment, multiple cameras can be installed in various locations including onboard Ole III. This approach may be computationally expensive but decrease the probability of misdetection errors.

In this research, YOLO v3 is used for the object detection task because of its relatively higher processing speed. In addition, practical applications of detection systems on autonomous vessels require real-time detection. This also justifies the selection of YOLO. Another recommendation for future research can be the implementation of the other object detection algorithms such as ResNet, SSD, RCNN to compare the detection accuracy and performance with the setting in this thesis.

7 References

Abilio Ramos, M., Utne, I. B., & Mosleh, A. (2019). Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events. *Safety Science, 116*, 33-44. doi:https://doi.org/10.1016/j.ssci.2019.02.038

AlexeyAB. (2019a). AlexeyAB/darknet. Retrieved from https://github.com/AlexeyAB/darknet

AlexeyAB. (2019b). AlexeyAB/Yolo_mark. Retrieved from https://github.com/AlexeyAB/Yolo_mark

Bekhet, A., & Zauszniewski, J. (2012). *Methodological triangulation: An approach to understanding data* (Vol. 20).

COCO. (2019). http://cocodataset.org/index.htm#home. Retrieved from http://cocodataset.org/index.htm#home

Di, W. (2018). *Deep learning essentials : your hands-on guide to the fundamentals of deep learning and neural network modeling* A. Bhardwaj & J. Wei (Eds.), *Deep learning essentials* Retrieved from https://learning.oreilly.com/library/view/deep-learning-essentials/9781785880360/?ar

DNV GL. (2019). The ReVolt. Retrieved from https://www.dnvgl.com/technology-innovation/revolt/index.html

DPA. (2018). The Norwegian Data Protection Authority (DPA). Retrieved from https://www.datatilsynet.no/en/

Elkins, L., Sellers, D., & Monach, W. R. (2010). The Autonomous Maritime Navigation (AMN) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles. *Journal of Field Robotics, 27*(6), 790-818. doi:doi:10.1002/rob.20367

EMSA. (2018). *ANNUAL OVERVIEW OF MARINE CASUALTIES AND INCIDENTS 2018*. Retrieved from Lisboa:

Finlay, S. (2017). *Artificial Intelligence and Machine Learning for Business : A No-Nonsense Guide to Data Driven Technologies*. Great Britain: Relativistic.

Frankfort-Nachmias, C. (2015). *Research methods in the social sciences* (8th ed. ed.). New York, NY: Worth publishers.

Goodfellow, I. (2017). *Deep learning*. Cambridge, Mass: MIT Press.

Google. (2019). Google Map. Retrieved from https://www.google.com/maps

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. Retrieved from

Hikvision. (2019). Hikvision DS-2TD2636-10. Retrieved from https://www.hikvision.com/en/Products/Thermal-Camera/Network-Bullet-Camera/Bi-spectrum/DS-2TD2636-10

Hui, J. (2018). Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. Retrieved from https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088

IMO. (2003). Convention on the International Regulations for Preventing Collisions at Sea (COLREG). London.

IMO. (2019). IMO takes first steps to address autonomous ships. Retrieved from http://www.imo.org/en/mediacentre/pressbriefings/pages/08-msc-99-mass-scoping.aspx

Kapur, S. (2017). *Computer vision with Python 3 : image classification, object detection, video processing, and more Computer vision with Python three* Retrieved from https://learning.oreilly.com/library/view/computer-vision-with/9781788299763/?ar

Kathuria, A. (2018). What's new in YOLO v3? Retrieved from https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b

Kim, K., Hong, S., Choi, B., & Kim, E. (2018). Probabilistic Ship Detection and Classification Using Deep Learning. *Applied Sciences, 8*(6). doi:10.3390/app8060936

Komianos, A. (2018). *The Autonomous Shipping Era. Operational, Regulatory, and Quality Challenges* (Vol. 12).

Kongsberg Maritime. (2019). Autonomous ship project, key facts about YARA Birkeland. Retrieved from https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/4B8113B707A50A4FC125811D00407045?OpenDocument
Kretschmann, L., Burmeister, H.-C., & Jahn, C. (2017). Analyzing the economic benefit of unmanned autonomous ships: An exploratory cost-comparison between an autonomous and a conventional bulk carrier. *Research in Transportation Business & Management, 25*, 76-86. doi:https://doi.org/10.1016/j.rtbm.2017.06.002
Kristiansen, S. (2005). *Maritime transportation : safety management and risk analysis*. Amsterdam: Elsevier/Butterworth-Heinemann.
Lloyd's Register. (2017). *LR Code for Unmanned Marine Systems*. Retrieved from
MUNIN. (2016). Welcome to the MUNIN Project web page. Retrieved from http://www.unmanned-ship.org/munin/
Navidi, W. (2006). *Statistics for engineers and scientists*. Boston: McGraw-Hill.
Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of Human Factors and Ergonomics Society, 39*(2), 230-253. doi:10.1518/001872097778543886
PASCAL. (2019). The PASCAL Visual Object Classes Homepage. Retrieved from http://host.robots.ox.ac.uk/pascal/VOC/
Redmon, J., & Farhadi, A. (2016). *YOLO9000: Better, Faster, Stronger*. Retrieved from https://arxiv.org/abs/1612.08242
Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. Retrieved from https://pjreddie.com/media/files/papers/YOLOv3.pdf
Redmon J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. Retrieved from https://arxiv.org/pdf/1506.02640.pdf
Ren, S., He, K., Girshick, R., & Sun, J. (2016). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. Retrieved from Cornell University:
Rodin, C. D., de Lima, L. N., Andrade, F., Haddad, D. B., Johansen, T. A., & Storvold, R. (2018). Object Classification in Thermal Images using Convolutional Neural Networks for Search and Rescue Missions with Unmanned Aerial Systems. doi:https://doi.org/10.1109/IJCNN.2018.8489465
Rolls-Royce. (2016). *Remote and Autonomous Ships The next steps*. Retrieved from https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf
Sharp, J. A., Peters, J., & Howard, K. (2002). *Management of a Student Research Project, The* (3rd ed. ed.): United Kingdom: Gower.
Stack Overflow. (2017). Intersection-over-union between two detections. Retrieved from https://stackoverflow.com/questions/28723670/intersection-over-union-between-two-detections
Thieme, C. A., Utne, I. B., & Haugen, S. (2018). Assessing ship risk model applicability to Marine Autonomous Surface Ships. *Ocean Engineering, 165*, 140-154. doi:https://doi.org/10.1016/j.oceaneng.2018.07.040
Uijlings, J. R. R., Sande, K. E. A. v. d., Gevers, T., & Smeulders, A. W. M. (2012). *Selective Search for Object Recognition*. Retrieved from IJCV:
USN. (2018). Applied Autonomy Systems summit. Retrieved from https://www.usn.no/aass/
Waterborne TP. (2011). *Strategic research agenda, Waterborne Transport & Operations Key for Europe's Development and Future*. Retrieved from
Winovich, N. (2019). Feedforward Networks. Retrieved from https://www.math.purdue.edu/~nwinovic/deep_learning.html#Feedforward_Networks
Woods, D. D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2017). *Behind Human Error* (2 ed.).

Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering & System Safety, 165*, 155-169. doi:https://doi.org/10.1016/j.ress.2017.03.029
Wróbel, K., Montewka, J., & Kujala, P. (2018). System-theoretic approach to safety of remotely-controlled merchant vessel. *Ocean Engineering, 152*, 334-345. doi:https://doi.org/10.1016/j.oceaneng.2018.01.020
Yang-Lang, C., Amare, A., Lena, C., Yi Chun, W., Chih-Yu, H., & Wei-Hong, L. (2019). Ship Detection Based on YOLOv2 for SAR Imagery. *Remote Sensing, 11*(7), 786. doi:10.3390/rs11070786

Appendix A

The observational report template used for observer onboard evaluation is presented below.

Dato	Klokkeslett	Observatør	Skipper

Retning	TRAFIKK VED START		
Husøy-Husvik ☐ Husvik - Husøy ☐		Styrbord	Babord
Totalt Antall passasjerer	Antall fartøyer med kryssende kurs		
Antall Barn (<16 år):	Av dette:		
	Seilbåter under seil		
Antall som trenger assistanse:	Seilbåter på motor		
	Kajakker/robåter		
Sykler:	Motoriserte fartøy		
	Nyttefartøy		

Eventuelle andre objekter i farvannet

SEILAS			
Avvik fra sjøveisreglene			
Regel	Beskrivelse av hendelse	Beskrivelse av skippers på Ole 3 handling	Annet

Skipper på Ole 3 handlet for å unngå hendelser			
Regel	Beskrivelse av hendelse	Hva gjorde skipper på Ole3	Annet

Skjedde det noen uhell på denne krysningen? Hvis Ja, så beskriv.

Sjøfartsregler		Eksempler på situasjoner:
15	Krysningssituasjoner	Ole 3 måtte ga fra seg sin forkjørsrett,
16	Handling av give-way vessel	Kryssende fartøy avvek fra sjøveisreglene
17	Handling av stand-on vessel	Kryssende fartøy gjorde «feil» i sin unnamanøver

Since the project was a Norwegian national project the documents were prepared in Norwegian language. The translation of the document can be found in below table.

ID	Date	Time	Observer	Captain	Direction (1 Husøy – Husvil 2 Husvil-Husøy)	Passengers			
						Total	Under 16 Years Old	Need Help	Bicycles

Traffic Situation									
Starboard					Port				
Crossing Vessels	Sailboat, Sailing	Sailboat, Motor	Kayak, Rowboat	Commercial Vessels	Crossing Vessels	Sailboat, Sailing	Sailboat, Motor	Kayak, Rowboat	Commercial Vessels

Deviations from COLREGS				How OLE III captain resolved the situation				
Regel	Description of incident	Description of Ole III captain´s action	Other	Rule	Description of incident	Description of Ole III captain´s action	Other objects in water	Note