

OIL SPILL DETECTION USING DEEP NEURAL NETWORKS

A Thesis

by

ZAHRA GHORBANI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Amir H. Behzadan
Co-Chair of Committee,	Phil Lewis
Committee Member,	Zhe Zhang
Head of Department,	Patrick Suermann

December 2019

Major Subject: Construction Management

Copyright 2019 Zahra Ghorbani

ABSTRACT

Oil spills have catastrophic effects on the environment, wildlife, economy, and human health. Therefore, timely detection of oil spills can reduce these disastrous impacts. Existing oil spill detection practices include in-situ (e.g., acoustic method, vapor sampling, pressure-point-analysis, and negative pressure wave) and remote sensing methods (e.g., traditional image processing and image processing using artificial intelligence). These methods rely mostly on skilled personnel for data collection, processing, and analysis, thus leading to slow, costly, and subjective results. Furthermore, oil platforms and pipelines are often situated in remote, harsh areas, making inspections hazardous. To remedy this problem, in this Thesis, three state-of-the-art artificial intelligence (AI) models, namely VGG16, YOLOv3 (you-only-look-once), and mask R-CNN (mask region-based convolutional neural network) are used in a transfer learning scheme to facilitate the process of detecting oil spills and surrounding objects such as vessels and oil rigs. Keyword search, a semi-supervised machine learning approach, is used to collect red-green-blue (R-G-B) imagery for training and testing these models. The methodology includes image classification, object detection, and instance segmentation. The VGG16 model is used to predict the existence of an oil spill in an image, yielding an accuracy of 93%. The YOLOv3 model is implemented to detect and mark the location of vessels and oil rigs. The mean average precision for detecting these two object classes is 61.5% (46% for vessel and 77% for oil rig). The mask R-CNN model is utilized to identify oil spill boundaries at the pixel level in the input image. Results (considering all test images)

indicate an average precision of 62%, and an average recall of 71%. Findings of this Thesis are sought to benefit oil and gas industry stakeholders and coastal communities by creating operational AI-assisted technologies for timely detection and response to oil spills and other environmental pollutions, ultimately contributing to human health, environment preservation, and profitability of energy exploration projects.

DEDICATION

I dedicate this Thesis to my parents, Maryam Jabbari and Nader Ghorbani. To my mother, I owe thanks for my unbreakable spirit and my will to persevere. I learned these strengths from watching her daily. To my father, I owe thanks for my confidence and my boldness. He taught me to always use my voice. Without their love and support, I would not be who I am today.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Amir H. Behzadan for his support and guidance. His level of professionalism, his field knowledge, and his organizational skills contributed to me finishing a thesis I am proud of. Our collaboration taught me academic skills as well as interpersonal communication skills in a professional setting. I would also like to thank my committee co-chair, Dr. Phil Lewis, for his calm reassurance along the way and my committee member, Dr. Zhe Zhang, for her attention to the details.

I would like to show my gratitude to the Construction Science department head, Dr. Patrick Suermann, for his endless support and kindness throughout my studies. He was always willing to meet and problem solve with me. I will leave Texas A&M University with the leadership skills I learned from him.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Amir H. Behzadan (Chair) and Professor Phil Lewis (Co-Chair) of the Construction Science Department, and Professor Zhe Zhang (Member) of the Department of Geography. All other work conducted for the Thesis was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from the Department of Construction Science at Texas A&M University.

NOMENCLATURE

AI	Artificial Intelligence
ANN	Artificial Neural Network
AveP	Average Precision
CNN	Convolutional Neural Network
COCO	Common Objects in Context
Coeff.	Coefficient
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSS	Decision Support System
EU	European Union
FC	Fully Connected
FN	False Negative
FP	False Positive
FPS	Frames per Second
GIS	Geographic Information System
GPU	Graphics Processing Unit
HPRC	High Performance Research Computing
HRQoL	Health-Related Quality of Life
IoU	Intersection over Union

mAP	Mean Average Precision
Max	Maximum
Min	Minimum
ML	Machine Learning
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NPW	Negative Pressure Wave
PAHs	Polycyclic Aromatic Hydrocarbons
PASCAL	Pattern Analysis, Statistical Modeling & Computational Learning
PPA	Pressure-Point-Analysis
PTSD	Post-Traumatic Stress Disorder
RGB	Red-Green-Blue
RNN	Recurrent Neural Network
RoI	Region of Interest
ROV	Remotely Operated Vehicle
SAR	Synthetic Aperture Radar
SD	Standard Deviation
TN	True Negative
TP	True Positive
UAV	Unmanned Aerial Vehicle
VOC	Visual Object Class

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGMENTS.....	v
CONTRIBUTORS AND FUNDING SOURCES.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES.....	xi
LIST OF TABLES	xiv
CHAPTER I INTRODUCTION	1
Oil Spill Impacts on the Economy, Human Health, Wildlife, and the Environment	4
Economic Impacts of Oil Spills.....	4
Human Health Impacts of Oil Spills	6
Wildlife and Environmental Impacts of Oil Spills	9
Research Objectives and Contributions	10
CHAPTER II OIL SPILL DETECTION AND RESPONSE.....	12
Oil Spill Response Process.....	12
Oil Spill Detection Techniques	13
In-Situ Methods.....	15
Acoustic Methods.....	15
Vapor Sampling.....	16
Pressure-Point-Analysis (PPA)	17
Negative Pressure Wave (NPW)	18
Remote Sensing Methods.....	19
Traditional Image Processing.....	19
Image Processing Using Artificial Intelligence.....	22
CHAPTER III DEEP LEARNING	24

Machine Learning	24
Supervised vs. Unsupervised Learning	24
Hyperparameters	25
Performance Metrics	27
Transfer Learning.....	30
Deep Learning.....	31
CHAPTER IV RESEARCH METHODOLOGY	37
Image Collection and Labeling	41
Dataset Description	41
Image Recognition	45
Image Classification	45
Training, Validation, and Testing.....	46
Image Classification Performance.....	47
Object Detection.....	49
Object Detection Performance	52
Model Retraining after Data Augmentation.....	54
Object Segmentation	56
Object Segmentation Performance	62
CHAPTER V CONCLUSIONS	72
Limitations	72
Future Work	74
REFERENCES	78

LIST OF FIGURES

	Page
Figure 1 Geographic Distribution of Thirteen Major Oil Spills around the Globe.....	2
Figure 2 Oil Spill Response Standard Protocol.....	13
Figure 3 Leak Detection and Localization Using Vapor Sensing Tube.....	17
Figure 4 Schematic Diagram of NPW Method for Leakage Detection.	19
Figure 5 A Framework for Oil Spill Detection Algorithms, Using SAR Images.	20
Figure 6 Supervised vs. Unsupervised ML Algorithms.....	25
Figure 7 Confusion Matrix for a Multiclass Classification.....	28
Figure 8 Confusion Matrix of a Binary-Class Classification.....	29
Figure 9 Learning Process in Traditional ML (Left), and Transfer Learning (Right).	31
Figure 10 Relationship between AI, ML, and DL.....	32
Figure 11 Traditional ML Algorithm vs. DL Algorithm.	32
Figure 12 Traditional ANN (Left) vs. DNN (Right).....	34
Figure 13 Object Recognition Tasks.....	35
Figure 14 Object Detection Methods.....	36
Figure 15 Schematic Diagram of the Methodology.....	38
Figure 16 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by YOLOv3 (Right).....	40
Figure 17 Sample Unseen Test Image (Left) and Corresponding Pixel-Level Segmentation by Mask R-CNN Model (Right).....	40
Figure 18 A Sample Image from the Dataset (Left) and Corresponding Segmented Image in LabelBox (Right).....	41
Figure 19 Distribution of the Number of Samples per Class Label.....	42

Figure 20 Distribution of the Number of Instances per Class Label.....	42
Figure 21 Weighted Venn Diagram for the Distribution of Number of Instances per Class Label.....	43
Figure 22 Examples from All Object Classes from the Dataset.	44
Figure 23 Examples of the Training Dataset Including Different Proportions of Oil Spill in the Image, Ranging from 5% to 90%.....	45
Figure 24 The Architecture of VGG16.	46
Figure 25 Training and Validation Accuracy per Epoch (Left) Training and Validation Loss (Right) (Training Step).....	48
Figure 26 Training and Validation Accuracy per Epoch (Left) Training and Validation Loss (Right) (Fine-Tuning Step).	48
Figure 27 Confusion Matrix for Oil Spill Classification.....	49
Figure 28 Key Steps in YOLO Implementation.....	50
Figure 29 Sample Image from the Test Dataset (Left) and Corresponding Detected Image by YOLOv3 (Right).....	51
Figure 30 Precision-Recall Curve for the Trained YOLOv3 Model.....	52
Figure 31 Distribution of the Number of Instances per Class Label, Before Data Augmentation (Left) and After Data Augmentation (Right).....	53
Figure 32 Example Image from the Training Set (Left) and Corresponding Augmented Images (Right).....	54
Figure 33 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by YOLOv3 (Right).	55
Figure 34 Precision-Recall Curve for the Retrained YOLOv3 Model.....	55
Figure 35 Mask R-CNN Algorithm.	56
Figure 36 R-CNN Algorithms.	57
Figure 37 Example of RoI Align.....	58
Figure 38 Transformation between Ground Truth and Prediction Bounding Boxes.	59

Figure 39 Loss Function per Epoch (a) Total (b) Classification (c) Bounding Box Regression (d) Mask.	60
Figure 40 Distribution of the Number of Oil Spill Images Based on Altitude.	61
Figure 41 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by Mask R-CNN (Right).	62
Figure 42 Positive and Negative Classes Based on the Detection and the Ground Truth.	63
Figure 43 Box Plots Summarizing Performance Metrics of Pixel-Level Segmentation of Test Images.	65
Figure 44 Sample Images from the Testing Set (Left) and Their Corresponding Detection Images by Mask R-CNN (Right).	68
Figure 45 Pixel-Level Performance Metrics of the Mask R-CNN Model Tested on Images from Different Altitudes.	70
Figure 46 Preprocessing of Images from the Testing Set to Marge Separate Masks into One Single Mask.	71
Figure 47 Example of False Positive Due to Low Contrast between Oil Spill and Background.	72
Figure 48 Example of False Positive Due to the Color of the Background.	74
Figure 49 Schematic Diagram of Triangulation.	76
Figure 50 Example of a Practical Scenario (From the Dataset Developed in this Thesis) Suitable for Triangulation and Mapping Purposes.	77

LIST OF TABLES

	Page
Table 1 Estimated Damages Caused by the Prestige Oil Spill (Garza-Gil et al., 2006). ...	5
Table 2 Major Oil Spills and Their Impacts on Human Health.	8
Table 3 Comparison between Leak Detection Methods (Murvay and Silea, 2012).	15
Table 4 SAR-Equipped Satellites (Brekke and Solberg, 2005).	20
Table 5 Overview of Deep Learning Models Used for Image Classification, Object Detection, and Instant Segmentation.	39
Table 6 VGG16 Model Summary.	46
Table 7 Selected Hyperparameters for the Classification Model.	47
Table 8 Hyperparameters Used for the Mask R-CNN Model.	61
Table 9 Performance Metrics Used for the Pixel-Level Evaluation on the Mask R- CNN Model.	64
Table 10 Measures of Central Tendency and Variability for Precision and Recall of Testing Set Images in Pixel Level.	67
Table 11 Performance Metrics of the Mask R-CNN Model Tested on Images from Different Altitudes.	69

CHAPTER I

INTRODUCTION

People from different backgrounds have quite similar perceptions of oil spills. The most common image that comes to mind when hearing the word “oil spill” is probably a large lost tanker that landed on a rock (Fingas, 2011). Although some major oil spills such as the Exxon Valdez incident (Palinkas, 2012), the *Hebei Spirit* spill (Sim et al., 2010), and the Prestige oil spill (Negro et al., 2009) have received more media coverage and public attention, the majority of people are not aware of the smaller scale spills that happen on a daily basis.

The importance of oil in all industry sectors is undeniable. With rising global oil consumption (Gately et al., 2012; Zou and Chau, 2006) especially in the developing regions of the world, oil pollution will also be on the rise (Bloch et al., 2015) which makes studying oil spills and their impacts on the environment, people, and wildlife even more crucial (Fingas, 2011). The author’s preliminary research has indicated that this global challenge affects many countries (both developed and developing) including the U.S., Canada, and China. Figure 1 shows thirteen major oil spills between 1967 and 2010: (1) the Persian Gulf incident (Kuwait, 1992); (2) the Gulf Oil Spill (Gulf of Mexico, 2010); (3) Ixtoc 1 Oil Spill (Mexico, 1979); (4) Atlantic Empress Oil Spill (Off the Coast of Trinidad and Tobago, 1979); (5) Kolva River Oil Spill (Kolva River, Russia, 1994); (6) Nowruz Oil Field Spill (Persian Gulf, Iran, 1983); (7) Castillo de Bellver Oil Spill (Saldanha Bay, South Africa, 1983); (8) Amoco Cadiz Oil Spill (Portsall, France, 1978);

(9) ABT Summer Oil Spill (Angola, 1991); (10) M/T Haven Tanker Oil Spill (Genoa, Italy, 1991); (11) Odyssey Oil Spill (Nova Scotia, Canada, 1988), (12) the Sea Star Oil Spill (Gulf of Oman, 1972); (13) the Torrey Canyon Oil Spill (Scilly Isles, U.K. , 1967) (Moss, 2010).



Figure 1 Geographic Distribution of Thirteen Major Oil Spills around the Globe.

The consequences of oil spills and slicks are catastrophic, and if not treated properly can negatively impact the environment, food chain, wildlife, and human safety in addition to producing a heavy financial burden on jurisdictions and governments. According to Rodriguez-Trigo (2007), oil spills can cause serious health effects on humans, such as damage to some genes or long-term respiratory devastating impacts. Beyer (2016) discussed some of the environmental effects of oil spills (in a case

study of the Deepwater Horizon oil spill). According to the findings of that study, species that were affected by the Deepwater Horizon oil spill included fish, deep-sea corals, sea turtles, and cetaceans. Therefore, finding an effective and fast way to detect oil spills could be a major first step toward reducing these disasters.

Traditional methods of oil spill detection include, among others, pressure-point-analysis (PPA) (Zhang, 1996), Otsu's method (Shu et al., 2010), and remote sensing (Brekke and Solberg, 2005). These methods require continuous involvement of skilled personnel in data collection, handling, and analysis that could be costly, slow, and more importantly, prone to subjectivity (based on one's best judgment of the problem parameters or solution space). In addition, in many cases, oil pipelines and platforms are located in remote and harsh areas, making it difficult and hazardous for engineers to conduct timely inspections.

Applying artificial intelligence (AI) and machine learning (ML) helps automate this process and create more objective measures by developing tools that can robustly learn from past disaster data and detect potential oil spills in new situations more precisely. This Thesis describes a scientific methodology designed and carried out by the author to enable the use of unmanned aerial vehicle (UAV) (a.k.a., drone) imagery along oil pipelines (on land) or offshore platforms (at sea). This imagery is used to train a deep learning model (a type of AI algorithm) on what an oil spill looks like as described by parameters such as color palette, pixel density, appearance, geometry, and progression pattern. Once trained on such data, the AI model can predict the likelihood of an oil spill or leakage in new footage obtained from other pipelines, oil rigs, or platforms.

To the best of author's knowledge, the majority of research on using aerial imagery for oil spill detection uses thermal images and Synthetic Aperture Radar (SAR). However, these methods are relatively expensive and significantly resource-intensive (Jha et al., 2008), may not be able to differentiate oil from some image backgrounds, such as soil and ice (Fingas and Brown, 1997), and depend on satellite imagery which may not be readily available for real-time applications (Girard-Ardhuin et al., 2003). Using ordinary drones to obtain Red-Green-Blue (RGB) images from the vicinity of oil spills could be potentially less expensive, more ubiquitous, and faster, making the oil spill response mission timely, cost-effective, and more practical. Existing literature in this area mostly reports on work conducted in Europe with a major focus on incidents that have occurred in geographical locations inside or around the European Union (EU) or the Middle East.

Oil Spill Impacts on the Economy, Human Health, Wildlife, and the Environment

Economic Impacts of Oil Spills

The extent of damages caused by oil spills on the economy is immense (Negro et al., 2009), which is evident by the increasing number of studies on oil spill effects on the economy over the past decade (Ritchie et al., 2013). Liu and Wirtz (2006) divided the cost imposed by oil spills into four main categories of socioeconomic losses, cleanup costs, environmental damages, and research costs. In another categorization, admissible claims are classified into three main groups, namely third-party claims, preventive measures, and natural resource damages (Loureiro et al., 2006). Negro et al. (2009) studied the economic impacts of the Prestige oil spill with a focus on the Death Coast fisheries and utilized fresh

fish landings to assess the amount of damage suffered. According to their findings, the weight of the species on the Death Coast dropped from 97.2% in 2001 to 84.5% in 2005.

Among other industries, tourism is the most vulnerable to disasters and crises. Ritchie et al. (2013) evaluated the short-term effects of the 2010 BP Gulf oil spill by measuring the performance of vacation rentals and hotel businesses. Results indicated that the BP oil spill generated both winners and losers depending on the industry type and geographical location. In particular, within the surrounding regions, hotel demand increased after the oil spill, specifically in the coastal communities of Alabama and Mississippi. One explanation could be that the hotel industry benefitted from accommodating the media and cleanup crews who chose to stay close to the area. Otherwise, a significant decrease in vacation rentals was reported in other locations.

Garza-Gil et al. (2006) investigated the short-term economic impacts of the Prestige oil spill on the Galician fishing and tourism industry. The estimated damages from their study are summarized in Table 1.

Table 1 Estimated Damages Caused by the Prestige Oil Spill (Garza-Gil et al., 2006).

Affected Sector	Loss (Million Euros)
Cleaning and Restoration	559.0
Coastal Fisheries and Aquaculture	64.9
Tourism	133.8
Total	761.7

Human Health Impacts of Oil Spills

When an oil spill occurs, many people are exposed to the chemicals and toxins, from residents who inhabit the area to those who are involved in the cleanup process. Oil spill toxins include various chemicals such as toluene, benzene, xylenes, ethylbenzene, polycyclic aromatic hydrocarbons (PAHs), vanadium, and mercury (Kadhim and Parry, 1984). Research shows that this exposure compromises both mental and physical health of humans (Aguilera et al., 2010; Levy and Nassetta, 2011).

Of the major oil spills around the world, the negative effect on human health has been investigated in only a few. Rodriguez-Trigo et al. (2007) studied human effects of the 2002 Prestige oil spill, and some acute symptoms such as headache, irritated eyes and throat, and respiratory symptoms were reported. Carrasco (2007) investigated the effects of the Prestige oil spill on health-related quality of life (HRQoL). Some of the health problems addressed in this study are social dysfunction, anxiety, insomnia, and depression. Palinkas (2012) studied psychological and social effects of the 2010 Deepwater Horizon oil spill and identified the most common mental health problems such as depressive symptoms, generalized anxiety disorders, Post-Traumatic Stress Disorder (PTSD), and intrusive stress.

Table 2 summarizes the findings of several studies related to the impacts of some major oil spills on human health. There is a consensus among these studies that exposure to oil spills can lead to severe endocrine, genotoxic, physical, and psychological impacts, in both short and long terms (Aguilera et al., 2010). Given the recurrence of these incidents, it is suggested to take proper actions to reduce these impacts. Goldstein et al.

(2011) emphasized the need to initiate some interference protocols, such as short- and long-term medical and psychological monitoring, particularly for more exposed populations such as oil spill cleanup workers.

Table 2 Major Oil Spills and Their Impacts on Human Health.

Study	Symptoms									
	Headache	Eye Irritation	Nausea	Dizziness	Throat Irritation	Fatigue	Skin Irritation	Respiratory Symptoms	Depression	Anxiety
1									×	×
2	×	×	×		×	×	×			
3	×	×			×				×	×
4	×	×	×		×	×				
5	×	×	×					×		
6	×	×	×	×				×		
7	×				×	×		×		
8	×	×	×							
9					×				×	
10	×	×	×	×	×	×	×		×	×
11	×	×	×	×			×	×		

(1) Exxon Valdez Oil Spill, Prince William Sound, AK, US (Palinkas, 1993) (2) Braer Oil Spill, Shetland, Scotland (Campbell et al., 1993) (3) Sea Empress Oil Spill, Milford Haven Waterway, UK (Lyons et al., 1999) (4) Nakhodka Oil Spill, The Sea of Japan (off the Oki Islands of Simane Prefecture) (Morita et al., 1999) (5) Erika Oil Spill, Brittany, France (Schvoerer et al., 2000) (6) Prestige Oil Spill, Galicia, Spain (Suarez et al., 2005) (7) Tasman Spirit Oil Spill, Karachi, Pakistan (Janjua et al., 2006) (8) Tasman Spirit Oil Spill, Karachi, Pakistan (Meo et al., 2009) (9) Hebei Spirit Oil Spill, Daesan port, South Korea (Song et al., 2009) (10) Hebei Spirit Oil Spill, Daesan port, South Korea (Lee et al., 2010) (11) Hebei Spirit Oil Spill, Daesan port, South Korea (Sim et al., 2010)

Wildlife and Environmental Impacts of Oil Spills

Regardless of their severity, the devastating impact of oil spills on wildlife and the environment is undeniable. Wildlife is directly affected by chemicals released during an oil spill. However, the damaging effects of small-scale oil spills are often overlooked since they lead to less immediate impacts in comparison to major oil spills (Cohen, 1995). Nonetheless, it is essential to study these damages to take measures to preserve wildlife. Romero and Wilkelski (2002) introduced an evaluation tool for monitoring the health status of endangered wildlife. In their study, marine iguanas (*Amblyrhynchus cristatus*) were observed for 20 years before a minor oil spill that occurred in Galapagos archipelago in 2001. Based on their findings, 62% of these species deceased one year after the oil spill. In investigating the reasons behind the death of these marine iguanas, the study found that oil residues killed their digestive bacteria, thus resulting in starvation.

Dubansky et al. (2013) studied the effects of the Deepwater Horizon oil spill on fish, particularly, Gulf killifish (*Fundulus grandis*). Results showed growing developmental abnormalities when Gulf killifish were exposed to oil residue. In another study, Swedmark et al. (1973) investigated the impact of oil dispersants on marine animals and suggested a sequential effect that included increased activity, successively impaired activity, immobilization, and death.

Oil spill pollution can remain in the environment for many years after the incident (Fingas, 2012). However, in most cases, if oil is removed from the scene, the recovery will be complete in as little as 2 to 10 years (Kingston, 2002). Therefore, it is necessary to

identify the impacts of oil spills on the environment and establish protocols for oil spill response.

Abii and Nwosu (2009) conducted a study on the impact of oil spills in two regions, Agbonchia and Ogali, in the Niger-Delta in Nigeria. According to the results, the oil spill had a significant impact on the fertility status and nutrient level of the soil. Beyer et al. (2016) found that the Deepwater Horizon oil spill had devastating biological impacts on the ecosystem. Some of the species affected by this incident were birds, marine mammals, pelagic fish, and sea turtles. The study suggested that certain species, such as large fish, sea turtles, cetaceans, and deep-sea corals should be monitored to control the long-term effects.

Research Objectives and Contributions

The main goal of this research is to enhance oil spill and leakage detection. Effective and rapid detection assists responsible authorities in taking proper actions to prevent further damage to the environment, economy, and human health. To achieve this goal, this study aims to meet the following primary objectives and answering specific questions related to each objective:

1. Study previous incidents from both environmental impacts and technical perspectives. To meet this objective, the following research question is answered:
 - a. What are the environmental, economic, ecological, and psychological impacts of oil spills, and their effects on flora, fauna, and humankind?

2. Explore existing datasets about oil spills and leakages. To meet this objective, the following research questions are answered:
 - a. Are existing datasets large and diverse enough to cover both land and sea incidents?
 - b. What are the characteristics and significant differences of these datasets in terms of quality and frequency when collected by satellites, planes, and drones?
3. Train AI models on oil spill visual data obtained from previous incidents, and test these models on data from other incidents to assess the applicability of the developed method to practical cases. To meet this objective, the following research questions are answered:
 - a. What level of accuracy can be achieved when the trained AI model is used to classify unseen images considering the general presence of oil spills in an image?
 - b. What level of accuracy can be achieved when the trained AI model is used to detect and localize (mark the locations) of oil spills and other related objects in unseen images?

This research is intended to introduce new AI-assisted tools that can lead to broader environmental, financial, and sociological impacts. It seeks to make a significant difference to the safety and wellbeing of people who depend on the petroleum industry from upstream (i.e., exploration and production), to midstream (i.e., transportation and storage), to downstream (i.e., refineries and distribution).

CHAPTER II
OIL SPILL DETECTION AND RESPONSE

Oil Spill Response Process

Oil spill response is a demanding task in any environment (Valez et al., 2011). Possible strategies to minimize oil spill damages can be divided into two categories, namely oil spill preparedness (including prevention) for operational and accident oil pollution, and control and recovery of a spill to mitigate its consequences (Ventikos, 2002).

Terms such as “cleanup” or “countermeasures” do not have their usual and literal sense. In the case of oil pollution, cleanup operations should be cautious, well-planned, and balanced because they can cause more damage than the pollution itself, e.g., due to the use of heavy machinery in sensitive spill sites, such as marshes. Therefore, it is important to compile and categorize an effective oil response plan to protect the environment (Ventikos et al., 2004).

Figure 2 shows the standard protocol that should be followed after an oil spill. First, oil spill should be detected using one or a combination of methods that will be described in this Chapter. This will be followed by notifying the appropriate authority who will implement proper actions to gain and maintain control, and prevent additional spread.

Next, cleanup operations will be carried out to avoid further impacts to the environment (Doerffer, 2013).

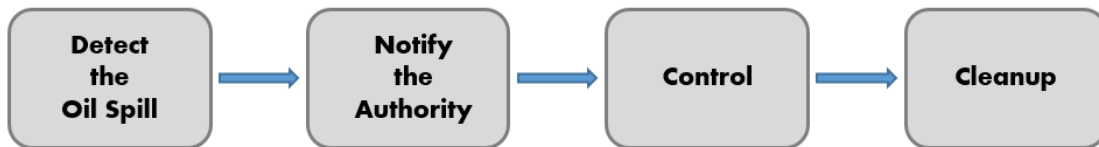


Figure 2 Oil Spill Response Standard Protocol.

Oil Spill Detection Techniques

In-situ (i.e., direct) oil spill detection methods often involve sensor installation or deployment of skilled personnel and resources to predetermined locations. To locate and detect oil spills using sensors, flow input and output parameters (e.g. pressure, discharge rate, velocity, and temperature) and physical properties of the system (e.g. pipe dimensions) must be known. Using this information, any divergence from normal state could be flagged as an indication of oil spill or leakage (Jiao et al., 2019).

Covas et al. (2005) classified these methods into direct observation and inference methods. Direct observation techniques assess pipeline features by means of specific tools such as infrared thermography, artificial patrol, video inspection, visual inspection, and acoustic techniques, with the latter considered as the most cost-effective direct observation method (Covas and Ramos, 1999). Inference oil spill detection methods can be used to monitor oil spills online. These methods rely on the pipeline features that are provided by other sensors or tools such as flow, pressure, and temperature. To detect oil spills, a model

that is based on real-time data must be established. Inference methods are mostly based on inverse analysis, a steady-state equation, and negative pressure wave (NPW) (Covas et al., 2005).

Zhang (1996) categorized oil spill detection methods into three groups, namely software-based, biological, and hardware-based methods. Software-based techniques include pressure-point-analysis (PPA), dynamic model system, mass or volume balance, and flow or pressure change. In biological methods, a leak is detected by a trained dog or by skilled personnel, using odor, visual symptoms, or sound. Results obtained by these methods rely mainly on human skills and may or may not be accurate (Murvay and Siela, 2012). Hardware-Based techniques utilize various hardware tools to detect oil spills, such as visual devices, pressure wave detectors, gas sampling devices, and acoustic devices (Zhang, 1996).

The limitation of in-situ oil spill detection methods lies in the fact that almost all of them require physical hardware (i.e., sensors, gauges, tools) installation, calibration, and maintenance. In addition, the majority of such methods is designed for onshore oil transit systems, such as pipelines. In this Chapter, a brief description of in-situ oil spill detection techniques is first presented. Next, follows an introduction to remote sensing techniques, such as those using traditional image processing and more recently, image processing using AI.

In-Situ Methods

In this Section, a brief description of some direct oil spill detection methods is presented that includes acoustic methods, vapor sampling, PPA, and NPW. Table 3 draws a comparison between these methods (Murvay and Silea, 2012).

Table 3 Comparison between Leak Detection Methods (Murvay and Silea, 2012).

	Method				
	Acoustic	Thermal imaging	Vapor sampling	PPA	NPW
Cost	H	H	H	L	L
Detection Speed	F	M	F	F	F
Easy Retrofitting	Y	-	N	Y	Y
Easy Usage	Y	Y	Y	Y	Y
Leak Localization	Y	Y	Y	N	Y
Leak Size Estimation	Y	Y	Y	N	Y

(Y)es, (N)o, (S)low, (M)edium, (F)ast, (L)ow, (H)igh

Acoustic Methods

When an oil leakage happens, noise is created as a result of fluid (oil) running from the pipeline. The velocity of the noise wave depends on the physical attributes of the fluid in the pipeline (Fuchs, 1991). In oil spill detection methods that use acoustic devices, specialized sensors detect these waves and generate a leak alert accordingly (Furness and Reet, 2009). It is recommended to use several detectors along the pipeline since the detection range of a single sensor is limited (Brodetsky and Savic, 1993). Installed sensors identify acoustic signals along the pipeline and distinguish the sounds caused by leaks from other sources of sound (Zhang, 1996). Examples of sensors that have been used for

the acoustic methods include microphones, dynamic pressure transducers, acoustic sensors, and accelerometers (Loth et al., 2003).

While the first uses of the acoustic methods date back to the 1930s (Rocha, 1989), the concept is still being used as part of more sophisticated alert systems (Loth et al., 2003). In a more recent study, a method was introduced based on the concept of acoustic methods in which time-frequency analyses were carried out to distinguish between background noises and the signals generated by the leakage (Meng et al., 2011). The main advantages of using these systems are the ability to conduct remote detection, and the relatively good performance of this technique to locate oil leaks (Murvay and Siela, 2012).

Vapor Sampling

Another method for detecting leaks involves sampling hydrocarbon vapors along the pipelines, using either a vapor monitoring system (Sperl, 1991) or mobile detectors (Bryce et al., 2002). A vapor monitoring system includes a sensor tube that is buried in the vicinity of the pipeline (Geiger et al., 2006). Figure 3 illustrates leak detection using vapor sensing tube. Mobile detectors can be mounted on remotely operated vehicles (ROVs) or be carried by skilled personnel (Bryce et al., 2002).

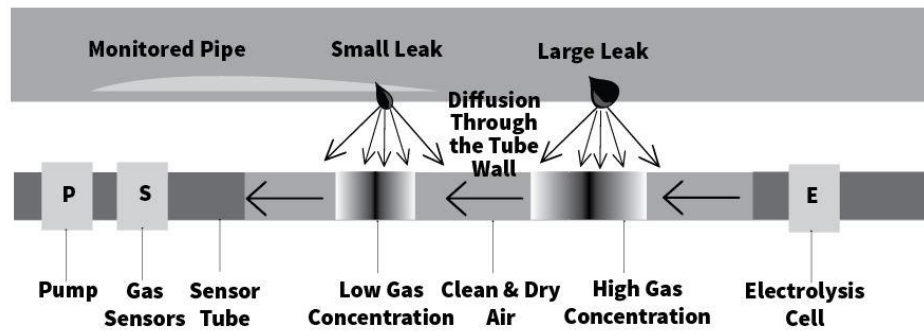


Figure 3 Leak Detection and Localization Using Vapor Sensing Tube.

The remote monitoring method works by using a sensor tube that is buried along the pipeline. This sensor tube can absorb the oil if a leak occurs. The content of the tube is analyzed periodically by pushing the tube content through a monitoring unit, using a pump. The size of the leak is then detected, based on the concentration profile, and the location of the leak is determined through comparing the travel time of the gas from a leak spot with the overall travel time (Geiger et al., 2006). Some of the limitations of this method, however, are that it is not applicable to deep, above-the-ground, or long pipelines, is costly, and has a slow response time (Bryce et al., 2002).

Pressure-Point-Analysis (PPA)

The concept behind PPA is that a drop in the line pressure indicates an oil leakage (Geiger, 2008). In this method, a drop in the pressure measurement mean value is identified through statistical analysis of pressure measurements. If the drop exceeds a predefined threshold, a leakage is reported (Zhang, 1996). This method has proven to detect leak rates less than 0.1% of flow (Murvay and Siela, 2012) in cold environments

and underwater (Scott and Barrufet, 2003). The main disadvantages of this method are that it requires continuous measurement of pressure along the pipeline and is not a reliable method for transient flow (Murvay and Siela, 2012).

Negative Pressure Wave (NPW)

When a leak occurs in a pipeline, there is a pressure drop in or around the leak location, which creates a rarefaction (i.e., negative pressure) wave in the pipeline (Silva et al, 1996). The speed of this wave, which moves both upstream and downstream from the location of the leak, equals the speed of sound in the pipeline (Zhang, 1996). This wave can, therefore, be measured by installing pressure transducers on both ends of every pipeline segment (Silva et al, 1996).

The oil leak can be located by comparing the time difference of the moments at which negative pressure wave occurs at each end of the pipe (Zhang, 1996). More recent systems that are based on negative pressure wave concept, such as Atmos wave, can estimate the leak size as well as locate it (Souza and Hoffman, 2011). The NPW method, however, is not practical for long-range pipelines (El-Sheikh, 2010). Figure 4 presents the schematic diagram of the NPW method.

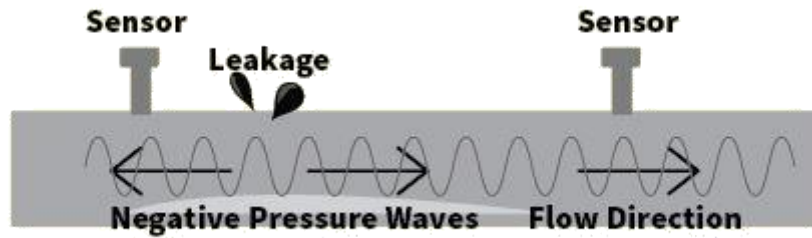


Figure 4 Schematic Diagram of NPW Method for Leakage Detection.

Remote Sensing Methods

Traditional methods of oil spill detection require continuous involvement of skilled personnel in data collection, processing, and analysis, which could be costly, slow, and more importantly, prone to subjectivity (based on one’s best judgment of the problem parameters or solution space). In addition, in many cases, oil pipelines and platforms are located in remote and harsh areas, making it difficult and hazardous for engineers to conduct safe, timely inspections. In light of these limitations, remote sensing oil spill detection techniques rely on information that can be obtained without a need for in-situ sensor installation and reading, using some of the approaches explained below.

Traditional Image Processing

Traditional image processing techniques rely on aerial data collected from the location of an oil spill. For instance, methods that use data captured by satellites equipped with synthetic aperture radar (SAR) have been previously proposed and tested in oil spill detection. SAR is a radar that is utilized to generate two-dimensional imagery or three-

dimensional objects (Kirscht and Rinke, 1998). Satellites with SAR can collect data at any time (day or night) and are independent of cloud coverage. These satellites operate at designated frequencies with L-band, S-band, and C-band wavelengths. Various SAR-equipped satellites are listed in Table 4 (Brekke and Solberg, 2005).

Table 4 SAR-Equipped Satellites (Brekke and Solberg, 2005).

Satellite (Sensor)	Operating Period	Owner	Band
SEASAT	1978	NASA	L-band
ALMAZ-1	1991-1992	Russian Space Agency	S-band
ERS-1	1991-1996	ESA	C-band
ERS-2	1995-present	ESA	C-band
RADARSAT-1	1995-present	CSA	C-band
ENVISAT (ASAR)	2002-present	ESA	C-band

L-band 1-2 GHz, S-band 2-4 GHz, C-band 4-8 GHz

In open bodies of water, oil leakages change the appearance of the water surface which in turn generates dark spots in satellite SAR imagery (Laur et al., 2002). Methods that use SAR imagery take advantage of these dark regions to detect oil spills (Curlandar and McDonough, 1991). Figure 5 presents the schematic framework for oil spill detection algorithms that use SAR images.

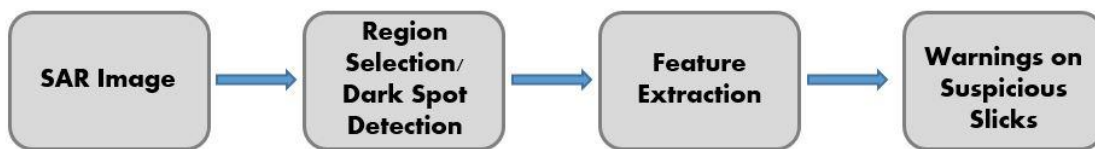


Figure 5 A Framework for Oil Spill Detection Algorithms, Using SAR Images.

Among common approaches to oil spill image processing is Otsu's method (Sezgin and Sankur, 2004; Zaart and Ghosn, 2013), which is an image thresholding technique at its core. One of the key advantages of this method is its efficiency and simplicity (Hou et al., 2006). The underlying algorithm uses a threshold to separate the pixels of the input image into two classes, foreground and background (Liao et al., 2001). To determine the threshold, the inter-class variance is maximized, or intra-class intensity is minimized (Otsu, 1979). The intra-class variance is defined as a weighted sum of variances of the two classes and is calculated using Equation 1 (Otsu, 1979). In this Equation, w_0 and w_1 represent the probabilities of the two classes that are separated by the threshold t , and σ^2_1 and σ^2_0 denote the variances of these classes.

$$\sigma^2_w(t) = w_0(t) \sigma^2_0(t) + w_1(t) \sigma^2_1(t) \quad (1)$$

It has been reported that Otsu's method performs poorly in some cases, including when there is a small mean difference between foreground and background pixels (i.e., low contrast), small object sizes, and in the presence of significant noise in the image (Lee and Chung, 1990).

Fiscella et al. (2000) proposed an approach that distinguished oil spills from other oceanographic phenomena using SAR imagery. Their method was based on two classifiers, namely Mahalanobis and compound probability, both of which measured the physical and geometrical features of the object and compared them with a template to determine if it was an oil spill. The classification accuracy was reported at 80%. Similarly, Solberg et al. (1999) introduced an approach that used adaptive thresholding technique to distinguish oil spills from other phenomena. Their method combined a statistical classifier

using a Gaussian model and several subclasses with a rule-based modification of prior probabilities and yielded a 94% accuracy.

In another study, Solberg et al. (2007) used two types of SAR imagery (i.e., RADARSAT and ENVISAT) for oil spill detection. Their method implemented a statistical classifier with subclasses based on wind and shape. The classification task considered features that described oil spill surroundings, oil spill and background contrast, shape, and homogeneity. Their method yielded a 78% accuracy. When benchmarked against manual oil spill detection, this algorithm was found to be faster with an average processing time of 1.45 minutes per image compared to 10 minutes per image in manual inspection.

Image Processing Using Artificial Intelligence

More recently, with advancements in processing capacities and big data analytics, methods that are built upon AI and fast image computing have gained traction in many fields. The introduction of AI and ubiquitous data acquisition platforms such as unmanned aerial vehicles (UAVs) and handheld devices with mobile connectivity in the oil and gas industry has started to change how oil spills and other types of environmental pollutions are detected and monitored.

Marquez et al. (2016) utilized drones equipped with various types of optical sensors to detect oil spills. In another study, Jiao et al. (2019) proposed an oil spill method that uses aerial imagery obtained from drones for oil spill detection. Their proposed approach consisted of four main sections including data preparation, model construction,

detection procedure, and responses to the results. In data preparation, UAVs were used to acquire raw training data. Next, several pre-trained DCNN models were used to detect the presence of oil spills in input images, and performance metrics (e.g., precision, recall) were used to evaluate the models. According to their findings, the DCNN achieved better results than the traditionally best method for oil spill detection, as measured by an overall accuracy of 98% and above. The study further reported that after one year of deployment in an oilfield, this method reduced inspection costs by 57% compared with the costs of the typical manual inspection, by reducing the number of workers and eliminating potential risks posed to workers by field conditions. However, the dataset used in this study was prepared using images taken in and around a particular location (city of Dongying, Shandong province, China) over a period of six months (March to December 2016), thus lacking diversity.

CHAPTER III

DEEP LEARNING

Machine Learning

Machine learning (ML), a subset of artificial intelligence (AI), uses algorithms to systematize the correlation between data and information (Awad and Khanna, 2015). ML has been succinctly defined as the “computer’s ability to learn something without being explicitly programmed” (Samuel, 1988). Mitchell (1997) offers a more formal definition of ML: “a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ”. ML algorithms have influenced everyday life in a variety of ways such as self-driving vehicles, web searches, and email filtering (Schmidhuber, 2015; Yu and Deng, 2011). They have been applied in many areas, including natural language processing (NLP) (Yu and Deng, 2011), information retrieval (Cheng and Baldi, 2006), and pattern recognition (Bishop, 2006).

Supervised vs. Unsupervised Learning

In a broader scheme, ML algorithms can be divided into two categories, namely supervised and unsupervised algorithms (Dunham, 2006). Figure 6 presents a schematic diagram of different ML algorithms. In supervised learning, the algorithm is built given both the input and output data. The input data (a.k.a., training data) includes training examples. Unsupervised learning, on the other hand, draws interfaces from the input data

without being provided with labeled responses (Russel and Norvig, 2010). Supervised algorithms are divided into two subsets of classification and regression (Alpaydin, 2010). In classification, the outputs are categorical values, such as activities or colors. Regression outputs, on the other hand, are numerical values within a range, such as height or temperature (Harrington, 2012).

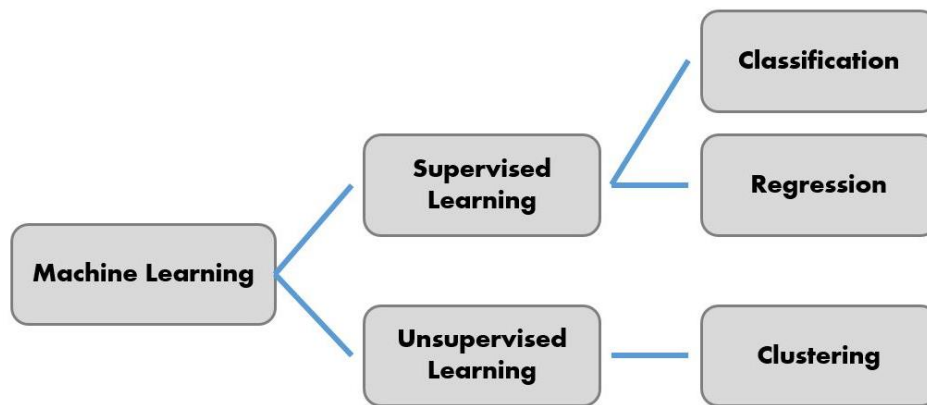


Figure 6 Supervised vs. Unsupervised ML Algorithms.

Hyperparameters

Hyperparameters are predefined parameters used by a ML algorithm to learn from training data (Claesen and Moor, 2015). Different ML algorithms use different hyperparameters (Claesen and Moor, 2015). In this research, examples of the hyperparameters that are used to train the ML models are learning rate, number of epochs, batch size, and intersection over union (IoU) threshold. These hyperparameters are briefly introduced below:

- *Learning rate*: defines how quickly the model updates its parameters during training.
- *Number of epochs*: is the number of times the entire training data is passed through the model during training.
- *Batch size*: indicates the number of training data samples in one forward or backward pass.
- *Number of steps*: a training step is one gradient update. During one step, several training data samples (as specified by the batch size) are processed.
- *IoU threshold*: for some ML tasks that deal with visual recognition (e.g., a computer vision algorithm to detect objects in an image), an IoU threshold is needed to differentiate between good and poor outputs, and determine the overall model performance in performing the desired task (classification or regression) (Cai and Vasconcelos, 2017). This hyperparameter is defined using Equation 2. In this Equation, area of overlap corresponds to the intersection of ground truth and detected area, and area of union is the union of ground truth and detected area.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (2)$$

- *Binarize threshold*: using the Binarize threshold t , an algorithm creates a binary image by replacing all values above t with 1 and remaining values with 0.

Performance Metrics

In this Subsection, a brief description of several mathematical approaches that are widely used to evaluate the performance of a ML model is provided.

Accuracy is perhaps the most intuitive metric used to describe the quality of the output generated by an ML model. Accuracy is defined as the ratio of number of instances that are correctly classified to the total number of instances (Dunham, 2006), as shown in Equation 3.

$$\text{Accuracy (\%)} = \frac{\text{No. of instances correctly classified}}{\text{Total no. of instances}} \times 100 \quad (3)$$

The accuracy for a binary (yes/no) classification is calculated by Equation 4. In this Equation, **True Positive (TP)** and **True Negative (TN)** refer to the number of correctly predicted values. TP indicates positive (i.e., yes) predictions that are correct, while TN indicates negative (i.e., no) predictions that are correct. **False Positive (FP)** and **False Negative (FN)**, on the other hand, refer to the number of incorrectly predicted values. FP indicates positive (i.e., yes) predictions that are incorrect, while FN indicates negative (i.e., no) predictions that are incorrect (Powers, 2011). Using these definitions, the denominator in Equation 4 is the total number of predictions (regardless of correctness) made by the model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

The performance of an ML model can be also presented in a matrix form known as the **confusion matrix**. Figure 7 shows an example of an n -by- n confusion matrix for a multiclass (n -class) classification task performed by a trained ML model. Rows in this matrix represent the actual (ground truth) classes, while columns indicate the predicted

classes. The (i, j) cell in the matrix indicates the number of instances that belong to class i but were predicted as belonging to class j (Harrington, 2012).

		Actual Value				
		Class 1	Class 2	Class 3	...	Class n
Predicted Value	Class 1	Correct				
	Class 2		Correct			
	Class 3			Correct		
	...				Correct	
	Class n					Correct

Figure 7 Confusion Matrix for a Multiclass Classification.

Knowing the number of TP, TN, FP, and FN instances, two more measures of performance can be calculated, namely *precision* and *recall*. Precision is defined as the number of TPs divided by the total number of predictions belonging to the positive class (i.e. TP+FP). Recall, on the other hand, refers to the number of TPs divided by the total number of TPs and FNs (Powers, 2011). Equations (5) and (6) are used to calculate precision and recall. Figure 8 presents a confusion matrix of a binary-class classification.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

		Actual Value	
		Positives	Negatives
Predicted Value	Positives	True Positive (TP)	False Positive (FP)
	Negatives	False Negative (FN)	True Negative (TN)

Figure 8 Confusion Matrix of a Binary-Class Classification.

The *F-score* is the harmonic mean of precision and recall, calculated using Equation (7). In this Equation, β is a positive real value that is chosen by the user depending on the specific type of problem the ML model is expected to solve. For $\beta = 1$, the general formula for $F(\beta)$ is simplified to the form shown in Equation (8), and the calculated value is referred to as the *F-1 score*.

$$F(\beta) = \frac{(1+\beta^2)*(\text{precision*recall})}{(\beta^2*\text{precision}+\text{recall})} \quad (7)$$

$$F(1) = 2 * \frac{(\text{precision*recall})}{(\text{precision}+\text{recall})} \quad (8)$$

If precision $p(r)$ is defined and plotted as a function of recall r , then the area under $p(r)$ curve is termed *average precision (AveP)*, as shown in Equation (9). In simple terms, *AveP* is the average value of $p(r)$ when r varies between 0 and 1 (Dunham, 2006).

$$AveP = \int_0^1 p(r) dr \quad (9)$$

In a multiclass classification task, the average of all *AveP* values (corresponding to different classes) is calculated using Equation (10) and referred to as the *mean average precision*, or *mAP*. In this Equation, c is the number of classes (Power, 2011).

$$mAP = \frac{\sum_{i=1}^c AveP(i)}{c} \quad (10)$$

Dice coefficient is a similarity coefficient that is used as a performance metric in image segmentation problems (Milletari et al., 2016). The original formula, denoted in Equation (11), could be applied to discrete data. In this Equation, X and Y represent two sets. When working with Boolean data, the dice coefficient can be written as Equation (12) (Thada and Jaglan, 2013).

$$\text{Dice Coefficient} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (11)$$

$$\text{Dice Coefficient} = \frac{2TP}{2TP + FP + FN} \quad (12)$$

Transfer Learning

For a ML model to be properly trained, there is a need for sufficient amount of training data. However, in many cases for reasons such as data scarcity, it is not possible to gather sufficient data from the problem domain. Generating a high-performance learner for a target domain (application of interest) by training it on data from a related source domain can help remedy this problem (Weiss et al., 2016). This process is often called *transfer learning* and aims at “storing knowledge gained while solving one problem and applying it to a different but related problem” (West et al., 2007). Figure 9 illustrates the difference between traditional ML and transfer learning approaches.

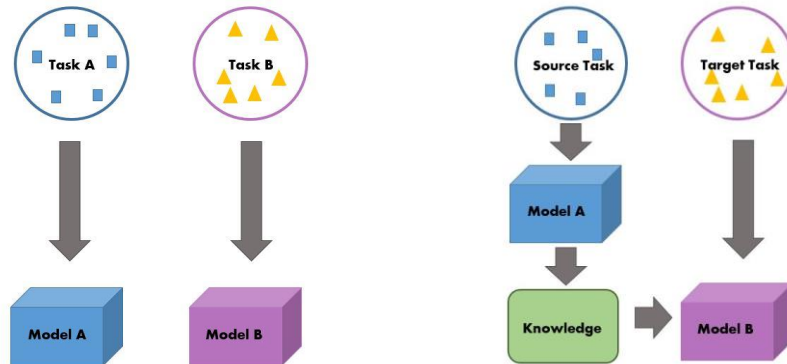


Figure 9 Learning Process in Traditional ML (Left), and Transfer Learning (Right).

Deep Learning

Traditional ML algorithms may underperform when dealing with large amounts of complex data, such as vision and speech for which hand-picking distinctive features may not be practical (Indiveri and Liu, 2015). The earlier notions of deep learning (DL) originally emerged in 1943, motivated by deep hierarchical structures of human speech perception (Liu et al., 2017). DL algorithms have improved significantly in the past few decades, particularly in areas such as computer vision, image analysis, natural language processing (NLP), speech recognition, and information retrieval (LeCun et al., 2015; Noda et al., 2015; Wu et al., 2016). Figure 10 illustrates the relationship between DL, ML, and AI. Figure 11 presents the difference between traditional ML algorithm and DL algorithm. As shown in this Figure, the traditional ML approach differs from DL is that ML algorithms need complex feature engineering (extraction, ranking, selection), and feature

extraction should be performed prior to passing the features to the ML algorithm. However, in DL, the data will be passed to the network directly (Bengio, 2009).

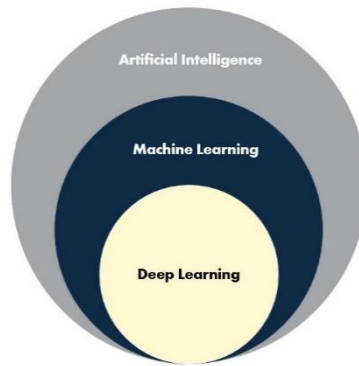


Figure 10 Relationship between AI, ML, and DL.

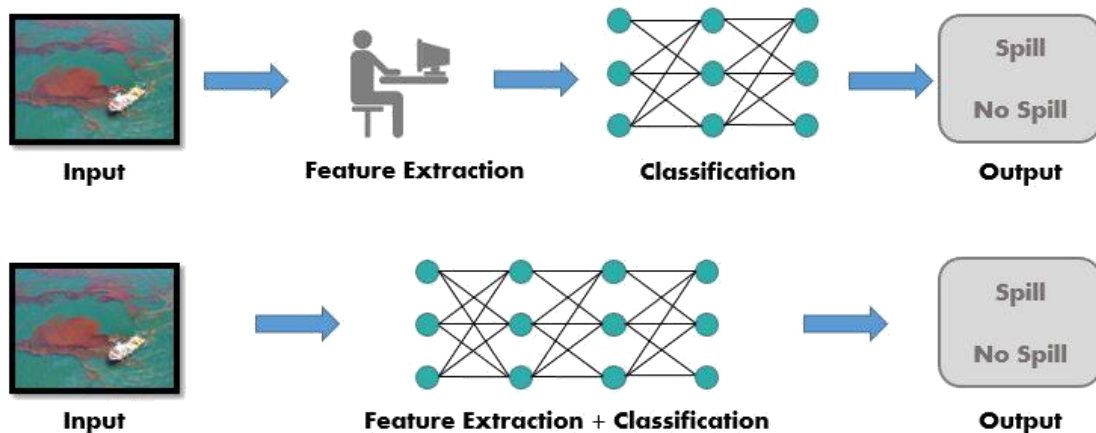


Figure 11 Traditional ML Algorithm vs. DL Algorithm.

(Images reprinted with permission from: Skimming Oil in the Gulf of Mexico during the Deepwater Horizon Oil Spill by Flickr user *Office of Response and Restoration*, CC-BY-2.0.)

DL architectures, such as artificial neural networks (ANNs), deep neural networks (DNNs), and recurrent neural networks (RNNs) (Hinton and Salakhutdinov, 2006) have

been applied in various fields, such as speech recognition (Hinton et al., 2012), computer vision (Szegedy et al., 2015; Nath and Behzadan, 2019), bioinformatics (Min et al., 2017), medical image analysis (Milletari et al., 2017), and drug design (Jing et al., 2018).

DL particularly became popular after significance advancements in speech recognition domain in 2006 (Hinton and Salakhutdinov, 2006; Hinton et al., 2012). Zhao et al. (2019) listed three main reasons for the increasing popularity of DL, as (1) high performance computing systems, such as graphics processing unit (GPU); (2) large-scale annotated datasets, including ImageNet (Deng et al., 2009), Pattern Analysis, Statistical Modeling and Computational Learning (PASCAL) Visual Object Class (VOC) (Everingham et al., 2010), and Common Objects in Context (COCO) (Lin et al., 2014); and (3) major improvements in network structures, such as AlexNet (Krizhevsky et al., 2012), Overfeat (Sermanet et al., 2013), VGG (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2015), and Resnet (He et al., 2016).

The architecture of a DL model can be best described using the traditional notion of ANNs (Hinton and Salakhutdinov, 2006). The resulting DL model is often referred to as a DNN. Figure 12 presents a schematic diagram of an ANN and a DNN. As shown in this Figure, a DNN is an ANN with multiple hidden layers between the input and output layers. Computations can be done faster and more parallelized, using DNNs (Yosinski et al., 2014).

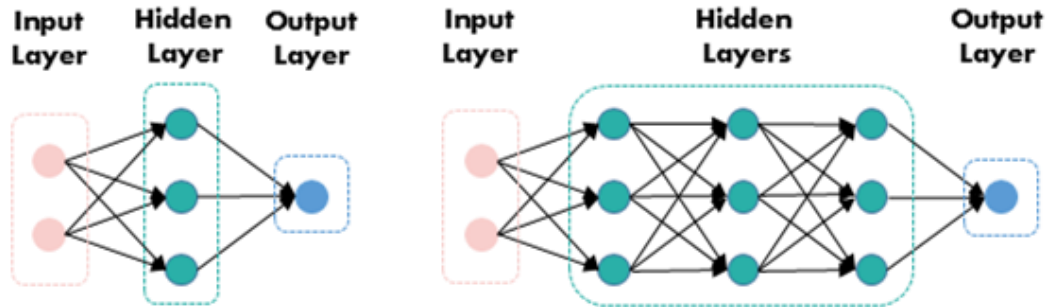


Figure 12 Traditional ANN (Left) vs. DNN (Right).

A convolutional neural network (CNN) is a type of deep learning model (LeCun et al., 2015). Each layer of a CNN encapsulates a feature map. The input layer feature map is a 3D matrix of pixel intensities for various color channels, such as RGB (Krizhevsky et al., 2012). Several types of transformation can be applied to feature maps, including pooling and filtering (Oquab et al., 2014). Pooling functions such as average pooling and max pooling, condense multiple cells (i.e., a field) of a feature map into one, and generate more vigorous feature descriptions (Kavukcuoglu et al., 2009). Filtering (i.e., convolution) function, on the other hand, convolutes a filter matrix (learned weights) with the values of a receptive field of neurons and takes a non-linear function to obtain final responses (Wadley, 1947). A feature hierarchy is created by interleaving between pooling and convolution. By adding various fully connected (FC) layers, with specific activation functions, this hierarchy can be fine-tuned and adapted for various tasks (Zhao et al., 2019).

In this Thesis, three well-established deep convolutional neural networks (DCNNs), including VGG16, you-only-look-once (YOLO), and mask region-based

convolutional neural network (R-CNN) are used for object recognition tasks. Object recognition, which is a field in computer vision, is used to identify objects in images or video frames (Torralba et al., 2003). Figure 13 outlines object recognition tasks, which fall into two main categories of image classification and object localization. In this Thesis, DCNN models are used for image classification (i.e., predicting the existence of an object in an image), object detection (i.e., detecting the location of an object in an image, using bounding boxes), and object segmentation (i.e., identifying the pixel-level boundaries of an object in an image). Figure 14 presents state-of-the-art object detection methods. From these methods, YOLO and Mask R-CNN are chosen for this research. These models are explained thoroughly in the following Chapter.

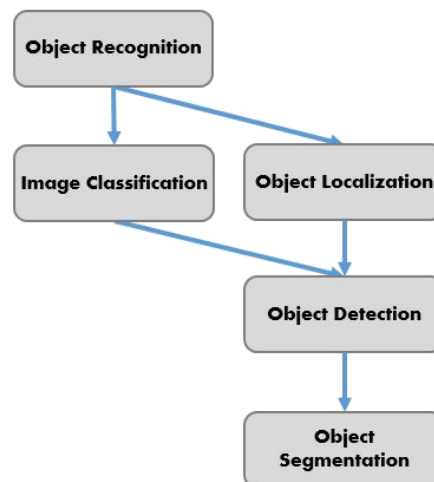


Figure 13 Object Recognition Tasks.

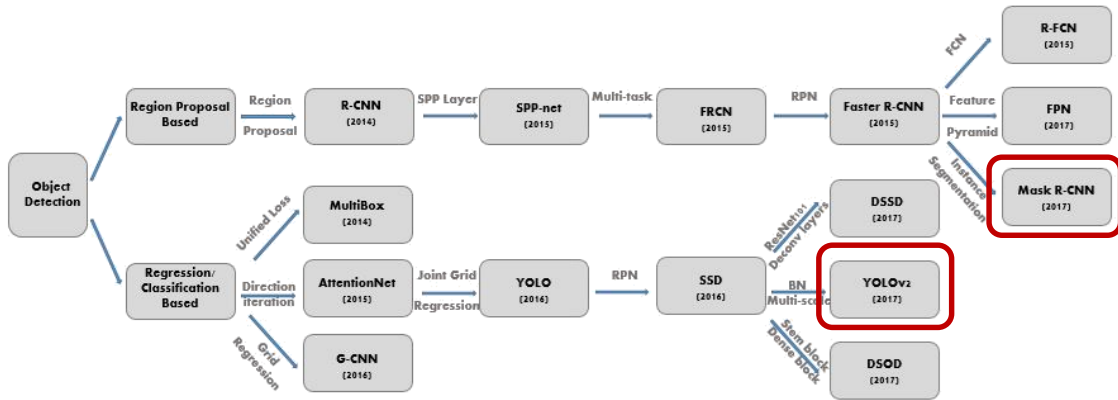


Figure 14 Object Detection Methods.

CHAPTER IV

RESEARCH METHODOLOGY

In this Chapter, the proposed solution to AI-assisted oil spill recognition in visual data is explained, which consists of image classification, object detection, and instance segmentation steps. Figure 15 presents the schematic diagram of the designed methodology. In data collection, relevant images are retrieved from Google using keyword search, a type of supervised web mining technique. Web mining is the process of integrating information gathered by traditional data mining techniques with those obtained from the world-wide web (Li, 2002). Following image retrieval through web mining, LabelBox (a web-based labeling toolbox) is used to label and annotate the visual data. Three object classes (oil spill, vessel, and rig) are annotated at the pixel-level (i.e., polygons are drawn around the object boundaries). Afterwards, the dataset is randomly split into three subsets of training (65% of the entire dataset), validation (15% of the entire dataset), and testing (remaining 20% of the entire dataset). This distribution follows common practices in machine learning where an 80-20 split is used to split the dataset into training and testing subsets, and the same ratio is applied within the training subset to separate initial training data from validation data (Nahato et al., 2015).

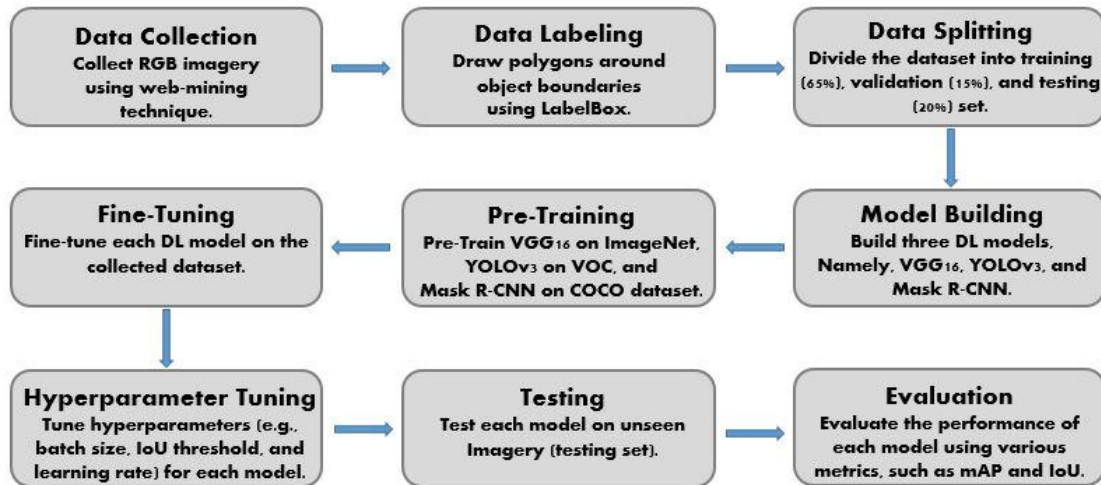


Figure 15 Schematic Diagram of the Methodology.

Next, three state-of-the-art deep learning models (pre-trained on general large-scale datasets) are selected. Table 5 lists these models, their network structures, and the datasets on which they are pre-trained. In particular, the VGG16 model is used for image classification. For object detection, YOLO (you-only-look-once) model is used to differentiate between two object classes (i.e., vessel and rig), and mask R-CNN (region-based convolutional neural network) is used for instance segmentation only for the oil spill class.

Table 5 Overview of Deep Learning Models Used for Image Classification, Object Detection, and Instant Segmentation.

Model	Network Structure	Pre-Trained Dataset
VGG16 (Krizhevsky et al., 2012)	16 Convolutional Layers	ImageNet (Deng et al., 2009)
YOLOv3 (Redmon et al., 2016)	Darknet-53 (Redmon and Farhadi, 2018)	VOC (Everingham et al., 2010)
Mask R-CNN (Gkioxari et al., 2018)	Resnet-50 (He et al., 2016)	COCO (Lin et al., 2014)

To implement the algorithms used in this research, Texas A&M University High Performance Research Computing (HPRC) clusters are used. In particular, the Terra cluster which comprises of an Intel x86-64 Linux cluster with 320 compute nodes with each node containing an Intel Xeon 2.5GHz E5-2670 v2 10-core processor is utilized (HPRC 2019). The total training time for the YOLOv3 and mask R-CNN models is approximately 5 hours and 18 hours, respectively.

Afterwards, pre-trained model weights are fine-tuned using the in-domain training images collected through web mining. In hyperparameter tuning, different hyperparameters such as batch size, learning rate, and the number of epochs is tuned for each model. This is followed by testing each model on unseen (i.e., test) images. Sample detections and their corresponding ground truth images are presented in Figures 16 and 17. Finally, the performance of the models is evaluated, using different evaluation metrics, including mean average precision, accuracy, recall, dice coefficient, and F1 score.

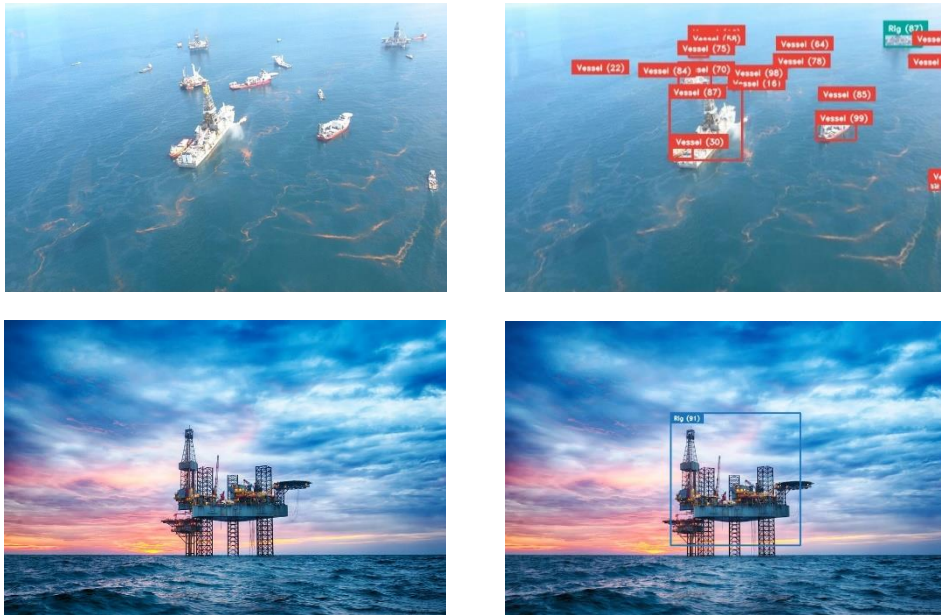


Figure 16 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by YOLOv3 (Right).

(Images reprinted with permission from: [View of Deepwater Horizon Oil Spill in May 2020](#) by Flickr user *Office of Response and Restoration*, CC-BY-2.0.; [HDR of Offshore Jack Up Rig in the Middle of the Sea at Sunset Time](#), Oil Rig by Flickr user *Solvay Group*.)

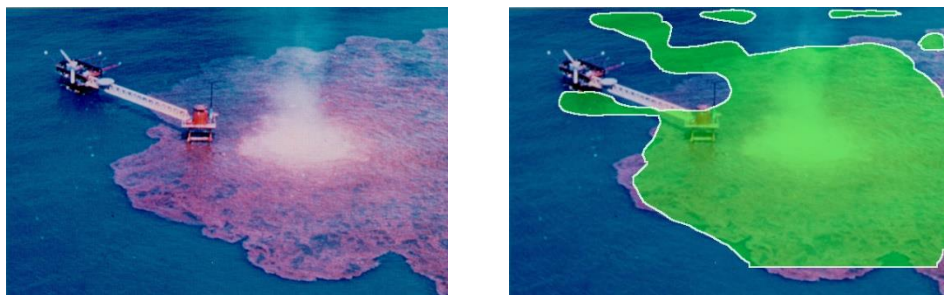


Figure 17 Sample Unseen Test Image (Left) and Corresponding Pixel-Level Segmentation by Mask R-CNN Model (Right).

(Image reprinted with permission from: [Ixtoc I Oil Spill](#) by Flickr user *NOAA Photo Library*.)

Image Collection and Labeling

As mentioned earlier, in order to create an image dataset in this research, relevant images are retrieved from Google using keyword search. Examples of keywords used are “oil spill”, “ocean aerial imagery”, and “sea aerial view”. Following image retrieval through web mining, LabelBox (a web-based labelling toolbox) is used to label and annotate the visual data. Figure 18 shows a sample image from the dataset and its corresponding image segmented in LabelBox.

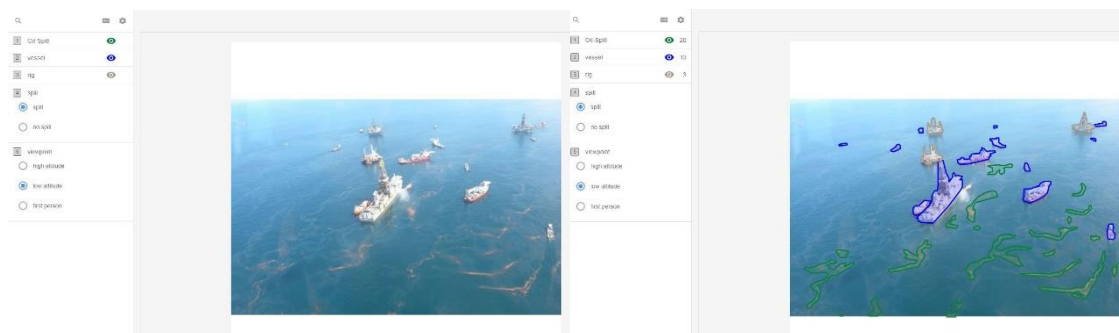


Figure 18 A Sample Image from the Dataset (Left) and Corresponding Segmented Image in LabelBox (Right).

(Images reprinted with permission from: [View of Deepwater Horizon Oil Spill in May 2020](#) by Flickr user *Office of Response and Restoration*, CC-BY-2.0.)

Dataset Description

The in-house dataset created in this research contains 1,292 images that are used to train, validate, and test several Deep Neural Network (DNN) models. The dataset is classified based on two criteria, namely the viewpoint (high altitude, low altitude, and first person) and the presence of oil spill (binary: yes/no). Figure 19 illustrates the distribution

of the dataset based on each criterion. In addition, each image is segmented by marking the boundaries of instances of three pre-defined object classes (i.e., oil spill, vessel, and rig). Each instance is marked by drawing polygons around the object boundaries. Figure 20 shows the distribution of the dataset based on the number of instances.



Figure 19 Distribution of the Number of Samples per Class Label.

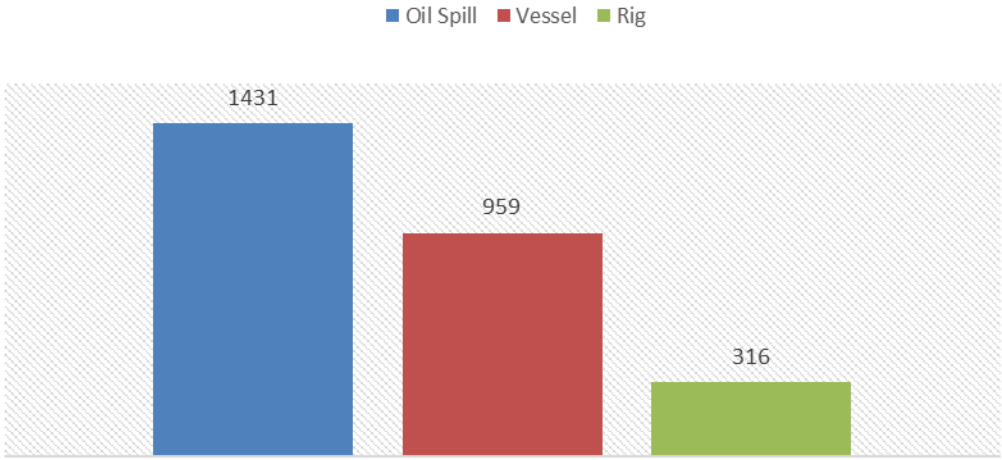


Figure 20 Distribution of the Number of Instances per Class Label.

Figure 21 shows the distribution of the dataset based on the instances in a weighted Venn diagram, and Figure 22 presents samples from the dataset.

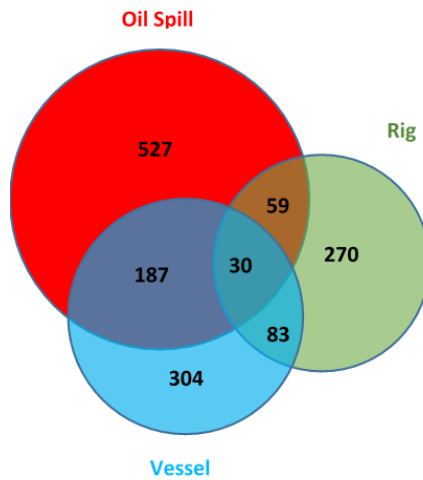


Figure 21 Weighted Venn Diagram for the Distribution of Number of Instances per Class Label.













LABELS	EXAMPLES			
Oil Spill				
Vessel				
Oil Rig				

Figure 22 Examples from All Object Classes from the Dataset.

(Images modified and reprinted with permission from: [Deepwater Horizon Oil Spill- May 24,2010](#) by *NASA/GSFC, MODIS Rapid Response*; [Ixtoc I Oil Spill](#) by Flickr user *NOAA Photo Library*; [Louisiana Oil Spill](#) by *US Coast Guard*; [Ixtoc I Oil Well Blowout](#) by *NOAA*; [Skimming Oil in the Gulf of Mexico during the Deepwater Horizon Oil Spill](#) by Flickr user *Office of Response and Restoration, CC-BY-2.0.*; [Deepwater Horizon Oil Spill- Gulf of Mexico](#) by *Kris Krug*; [Controlled Burn of Oil on May 19th](#) by *John Kepsimelis*; [Development-Driller-2](#) by *Barry Bena*; [HDR of Offshore Jack Up Rig in the Middle of the Sea at Sunset Time, Oil Rig](#) by Flickr user *Solvay Group*; [Oil Rig in the Gulf of Mexico](#) by Flickr user *Office of Response and Restoration*; [Off Shore](#) by Flickr user *arbyreed*; [Oil Platform in the Santa Barbara Channel, California 2](#) by Flickr user *Ken Lund, CC-BY-SA-2.0.*)

The area covered by oil spill in the training set ranges from 5% to 90%. Figure 23 depicts examples of the training set with different levels of oil spill coverage.



Figure 23 Examples of the Training Dataset Including Different Proportions of Oil Spill in the Image, Ranging from 5% to 90%.

(Images modified and reprinted with permission from: [Oil Leak from Damaged Well in Gulf of Mexico April 25th View](#) by Flickr user *NASA Goddard Space Flight Center*, CC-BY-2.0; [Striped Dolphins](#) by *NOAA's National Ocean Service*; [Montara Oil Spill- August 25, 2009](#) by Flickr user *SkyTruth*, CC-BY-NC-SA-2.0.; [Deepwater Horizon Oil Spill- Gulf of Mexico](#) by *Kris Krug*; [Deepwater Horizon Oil Spill Skimming Operations 2010-04-28](#) by *PO2 Prentice Danner*.)

Image Recognition

Image Classification

For image classification, a well-established deep learning network, namely VGG16 (Krizhevsky et al., 2012), is adopted and fine-tuned (through transfer learning). This network takes an RGB image as input, generates intermediate features through a series of convolution and max-pooling operations, passes the features to the fully-connected layer, and outputs the probabilities of the image belonging to any of the two classes of “spill” or “no spill”. As shown in Figure 24, the model is comprised of one input layer, 18 VGG16 layers, two FC layers, and one output layer. The VGG16 layers consist of a series of convolutional and max-pooling layers with 14,714,688 pre-trained weights. The output of the last VGG16 layer is connected to a flattened layer including 8,192 nodes,

which is fully connected to the next layer of 256 nodes. In this layer, a dropout operation is performed with 50% probability, i.e., during each iteration of the training session, 50% of the nodes are randomly excluded from weight updating. Together, the two FC layers contain 2,097,408 (i.e., 8192×256) weights. Table 6 summarizes the properties of the model.

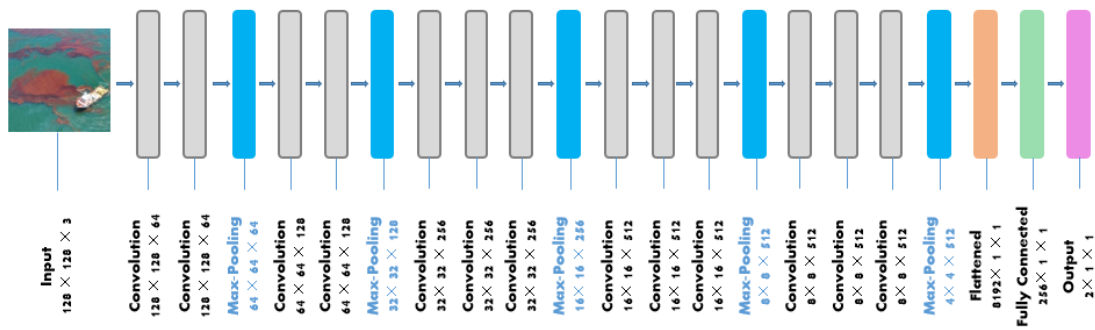


Figure 24 The Architecture of VGG16.

(Image reprinted with permission from: Skimming Oil in the Gulf of Mexico during the Deepwater Horizon Oil Spill by Flickr user *Office of Response and Restoration*, CC-BY-2.0.)

Table 6 VGG16 Model Summary.

Layer (Type)	Output Shape	Parameter #
VGG16 (Pre-trained Model)	(None, 4, 4, 512)	14714688
Flatten_1 (Flatten)	(None, 8192)	0
Dense_1 (Dense)	(None, 256)	2097408
Dense_2 (Dense)	(None, 1)	257

Training, Validation, and Testing

The training process consists of two steps. In the first step, which is referred to as “training”, the DNN model learns how to classify new images using the filters from the

pre-trained dataset. In this step, only the weights of the FC layers are updated, and the weights of the VGG-16 layers are frozen (i.e., not updated). The next step of training is referred to as “fine-tuning” during which all the previously frozen layers adapt to the new dataset with no significant change in their weights (Simonyan and Zisserman, 2014). In this step, the model is fed with the training dataset, and weight values of the last three convolutional layers and two FC layers are updated again. The hyperparameters listed in Table 7 are selected empirically. Finally, for model validation, the validation dataset was tested on the trained model.

Table 7 Selected Hyperparameters for the Classification Model.

Hyperparameter	Value
Epoch	30
Batch size	20
Learning rate	10^{-4}

Image Classification Performance

In order to evaluate model performance in correctly classified images based on the presence of oil spill, precision, accuracy, and recall metrics are used. As shown in Figures 25 and 26, in both training and fine-tuning steps, the training loss decreases with each epoch, while the training accuracy increases with every epoch (Chollet, 2018).

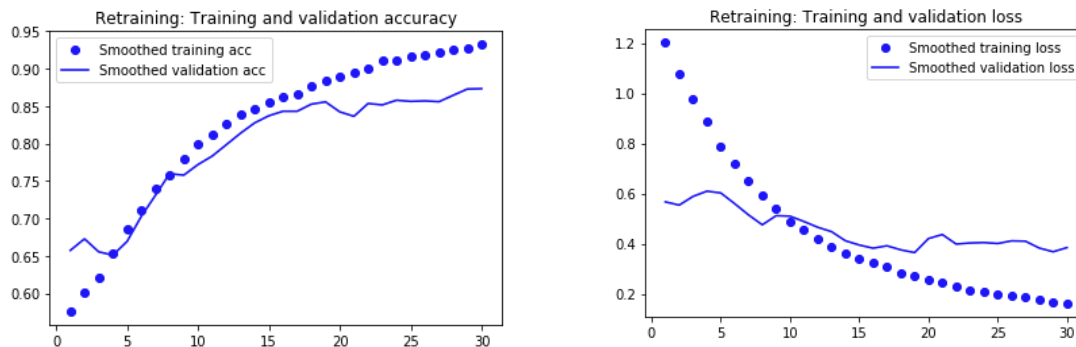


Figure 25 Training and Validation Accuracy per Epoch (Left) Training and Validation Loss (Right) (Training Step).

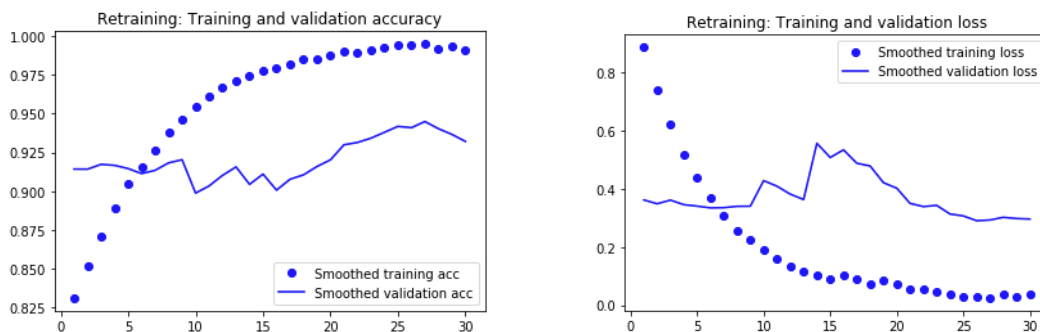


Figure 26 Training and Validation Accuracy per Epoch (Left) Training and Validation Loss (Right) (Fine-Tuning Step).

The model achieves an overall accuracy of %92.77 and a test loss of %23.71. Classification results in form of a confusion matrix are shown in Figure 27, which indicates that in 83 test images, oil spill is classified correctly, while in only one of the test images it is classified incorrectly (i.e., the model predicts an oil spill while there is no oil spill in the image). Similarly, in 90 test images, no spill is classified correctly, while in 10

test images it is classified incorrectly (i.e. the model predicts that the image does not contain an oil spill while there is an oil spill in the image).

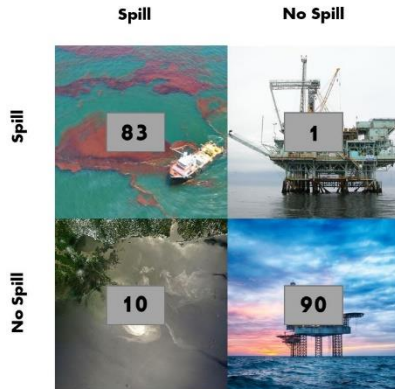


Figure 27 Confusion Matrix for Oil Spill Classification.

(Images reprinted with permission from: Skimming Oil in the Gulf of Mexico during the Deepwater Horizon Oil Spill by Flickr user *Office of Response and Restoration, Platform Hillhouse, Dos Cuadras (10)* by *Doc Searls*; CC-BY-2.0; Deepwater Horizon Oil Spill-May 24,2010 by *NASA/GSFC, MODIS Rapid Response*; HDR of Offshore Jack Up Rig in the Middle of the Sea at Sunset Time, Oil Rig by Flickr user *Solvay Group*.)

Object Detection

Object detection aims at localizing (i.e., marking the location of) three types of objects in an input image. In this research, these object types include “oil spill”, “vessel”, and “oil rig”. For this purpose, a state-of-the-art deep learning model is used, namely YOLO (You-Only-Look-Once) (Redmon et al., 2016). YOLO can predict bounding boxes and confidence scores for multiple categories by using the uppermost feature map. Figure 28 presents key steps of YOLO implementation. The process involves dividing the input

image into an $S \times S$ grid, and predicting bounding boxes and their corresponding confidence scores inside each grid cell.

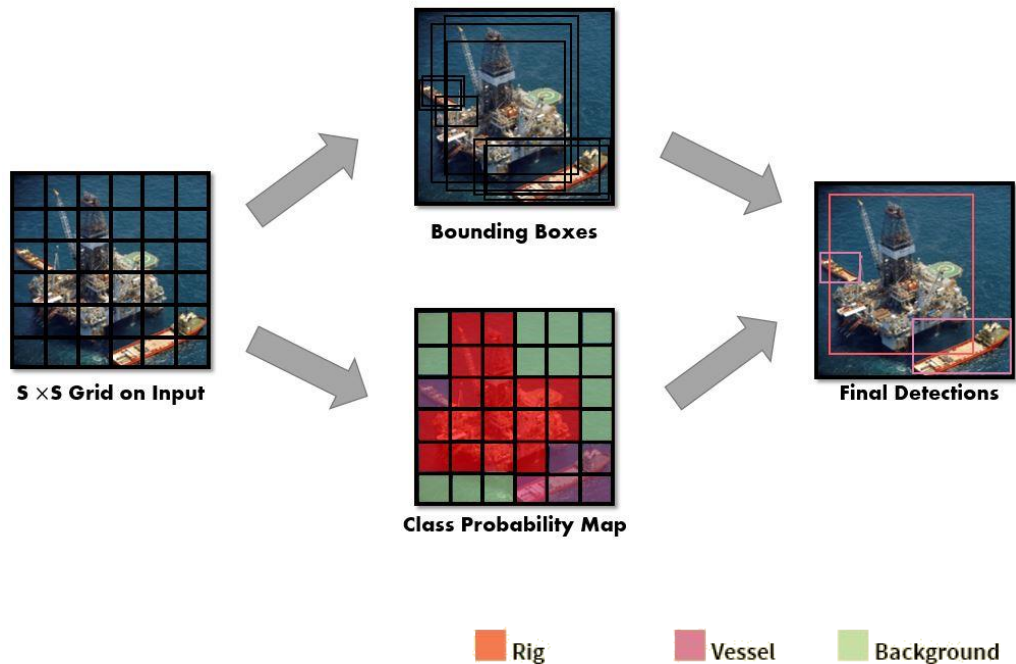


Figure 28 Key Steps in YOLO Implementation.

(Images reprinted with permission from: [Development-Driller-2](#) by *Barry Bena.*)

The YOLO architecture, established in 2017, includes two FC layers and 24 convolutional layers, some of which create ensembles of inception modules with 1×1 reduction layers, followed by 3×3 convolutional layers. This architecture can do real-time image processing at 45 frames per second (FPS) (Redmon et al., 17). Later, an enhanced version of YOLO, YOLOv2, was introduced, that implemented various outstanding strategies, including anchor boxes, and multi-scale training (Redmon and Farhadi, 2016). Afterwards, YOLOv3 was presented in 2018, with some updates to YOLOv2. YOLOv3 is capable of processing images at 30 FPS on a Pascal Titan X, and it has a mAP of 57.9%

on COCO test set. It can detect small objects significantly better than the previous models and is also faster when the speed of the detection is of high importance. Figure 29 presents a sample ground truth image from the test dataset, and its corresponding detections marked with bounding boxes.



Figure 29 Sample Image from the Test Dataset (Left) and Corresponding Detected Image by YOLOv3 (Right).

(Image reprinted with permission from: [Ixtoc I Oil Well Blowout](#) by NOAA.)

The mean average precision (mAP) for all three classes (oil spill, vessel, and rig) is 37.9%. Precision-recall curves and the average precision for each class are plotted in Figure 30.

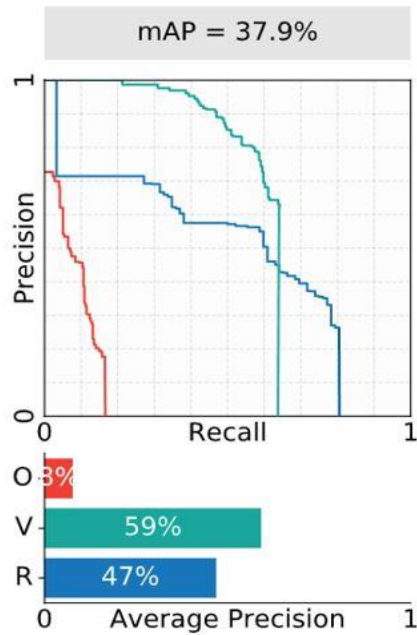


Figure 30 Precision-Recall Curve for the Trained YOLOv3 Model.

Object Detection Performance

The YOLOv3 model is initially used in this research for detecting all three object classes (i.e., oil spill, vessel, and rig). However, as shown in Figure 30, the average precision for the oil spill class is very low (8%) in comparison to other classes (vessel: 59% and rig: 47%). This could be potentially attributed to fuzzy and unclear borders of oil spills and their concave and arbitrary shapes (Chang et al., 2009). Therefore, the YOLO model is subsequently trained only on the other two classes (i.e., vessel and rig) and for the oil spill class, another DL model, mask R-CNN, is used for object detection and instance segmentation. As shown in Figure 31 (Left), the dataset used for retraining the YOLO model on vessel and rig classes is not balanced. To balance the dataset, data

augmentation is conducted on the rig class as the number of instances belonging to this class are significantly lower than the number of instances belonging to the vessel class.

Data augmentation is done as a preprocessing step by performing one or a combination of multiple operations on the initial image. Examples include flipping (both horizontally and vertically), rotating in different angles, and changing the blurriness, brightness, contrast, and hue and saturation (Chollet, 2018). In this research, changes in the blurriness of the images, hue and saturation, brightness, and rotation are used to increase the number of images that contain the minority class (i.e., rig). Figure 32(Left) presents an example image from the training set, and Figure 32(Right) presents resulting images after augmentation. As shown in Figure 31(Right), data augmentation results in a fairly balanced dataset by adding 564 new images that contain instances of class rig.

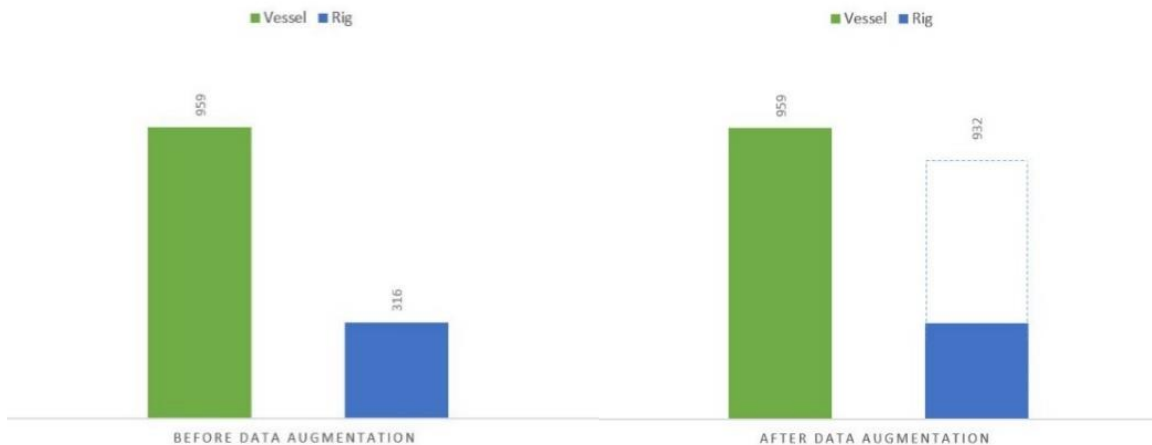


Figure 31 Distribution of the Number of Instances per Class Label, Before Data Augmentation (Left) and After Data Augmentation (Right).



Figure 32 Example Image from the Training Set (Left) and Corresponding Augmented Images (Right).

(Images modified and reprinted with permission from: [HDR of Offshore Jack Up Rig in the Middle of the Sea at Sunset Time, Oil Rig](#) by Flickr user *Solvay Group*.)

Model Retraining after Data Augmentation

Following data augmentation, a new YOLO model is trained on vessel and rig classes only. Figure 33 presents sample ground truth images from the test dataset, and their corresponding detections marked with bounding boxes. The mAP for the two classes (vessel and rig) is 61.5%. Precision-recall curves and the average precision for each class are plotted in Figure 34.

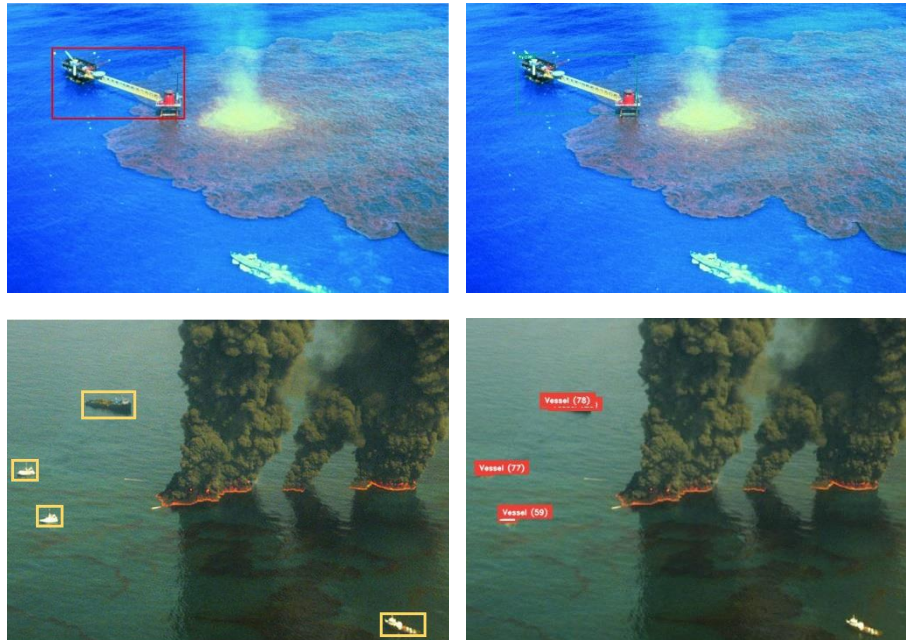


Figure 33 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by YOLOv3 (Right).
 (Images reprinted with permission from: Ixtoc I Oil Spill by Flickr user *NOAA Photo Library*; Controlled Burn of Oil on May 19th by *John Kepsimelis*.)

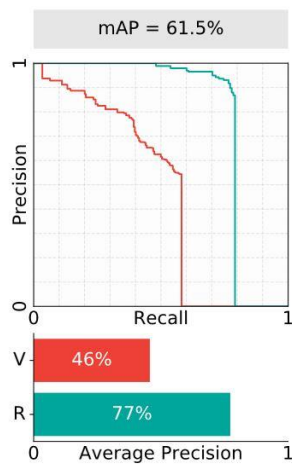


Figure 34 Precision-Recall Curve for the Retrained YOLOv3 Model.

Object Segmentation

The mask R-CNN model is used in this research to detect and segment instances belonging to the oil spill class. This model uses the mask R-CNN algorithm (Gkioxari et al., 2018), presented in Figure 35, for instance segmentation.

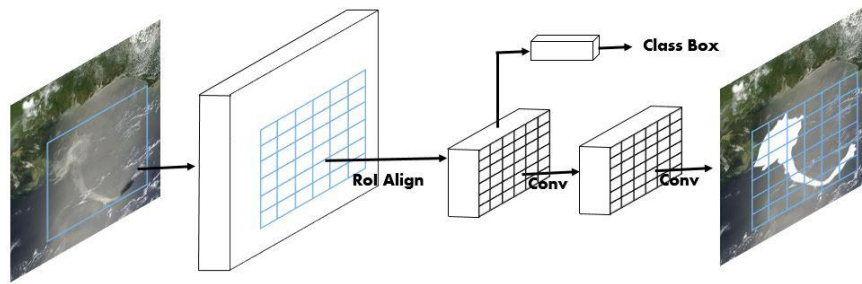


Figure 35 Mask R-CNN Algorithm.

(Image reprinted with permission from: Gulf of Mexico Oil Spill by *Rob Gutro*.)

Figure 36 illustrates the algorithms that are based on region-based convolutional neural networks, namely, R-CNN, fast R-CNN, faster R-CNN, and mask R-CNN. In this Thesis, mask R-CNN algorithm is used for instance segmentation. Instance segmentation comprises of detecting objects in an image and segmenting each object (i.e., semantic segmentation) (Arnab and Torr, 2017). These two tasks (i.e., object detection and semantic segmentation) are considered independent from one another, and prior to the introduction of mask R-CNN it was not possible to do both simultaneously. To remedy this issue, mask R-CNN uses the faster R-CNN algorithm (which does the classification and bounding box prediction) while also adding a branch to do pixel-level segmentation (He et al., 2017).

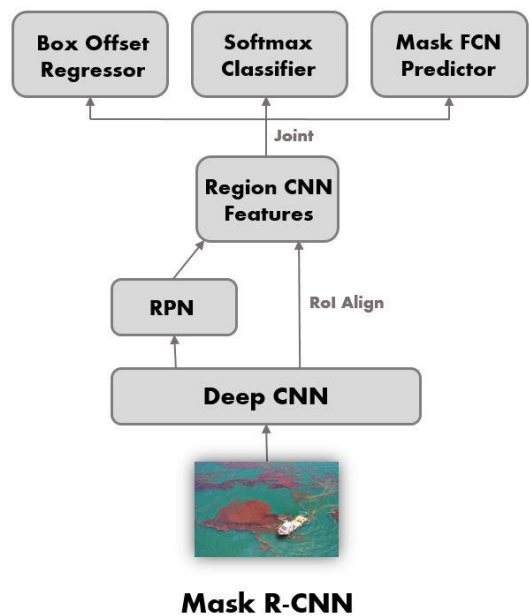
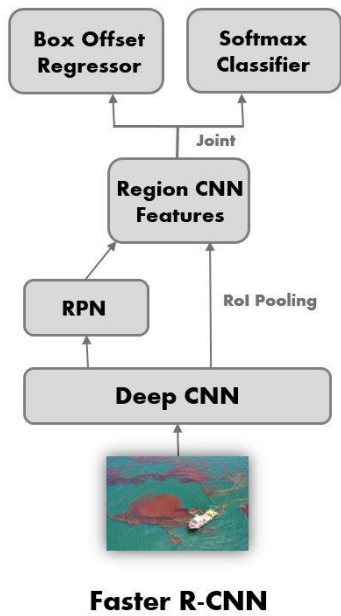
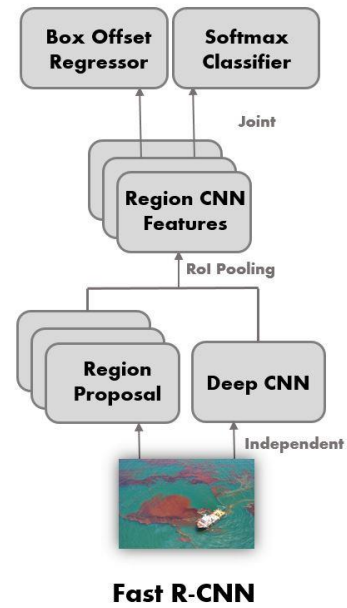
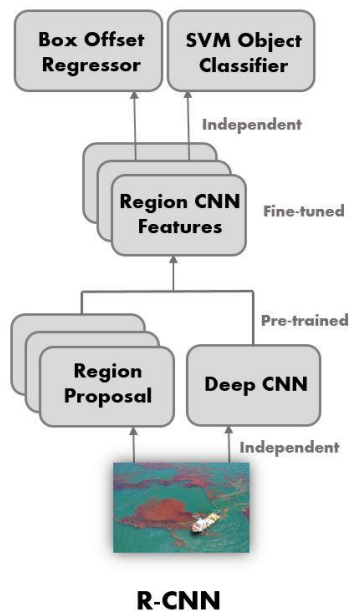


Figure 36 R-CNN Algorithms.

(Images reprinted with permission from: Skimming Oil in the Gulf of Mexico during the Deepwater Horizon Oil Spill by Flickr user *Office of Response and Restoration*, CC-BY-2.0.)

Mask R-CNN uses a layer, called region of interest (RoI) align, that implements bilinear interpolation (Jaderberg et al., 2015) to compute the precise values of the input features at four sampling locations in each RoI bin. This layer preserves the pixel-level spatial correspondence thus enhancing the mask accuracy (Zhao et al., 2019). Using this layer, an RoI can be mapped from the input image into the feature map accurately. Figure 37 illustrates example of the RoI align.

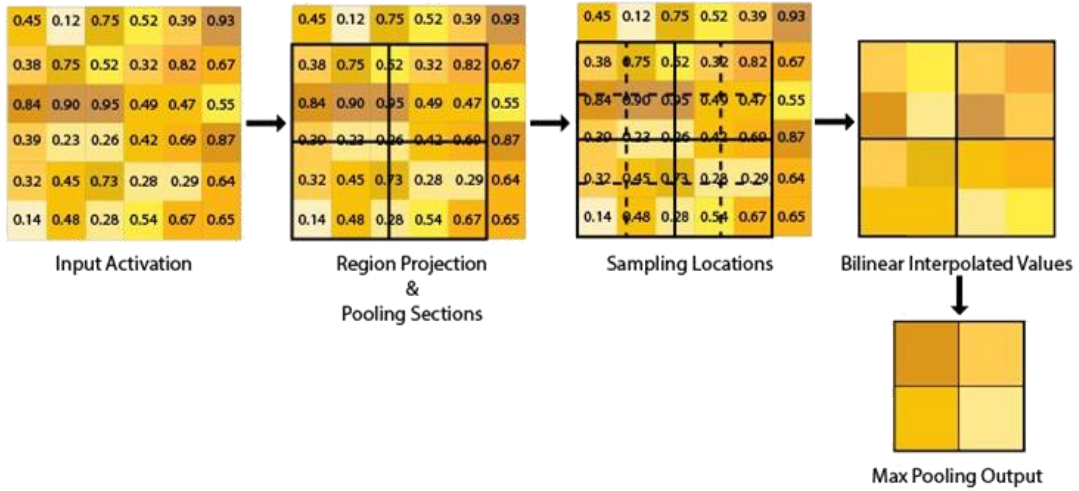


Figure 37 Example of RoI Align.

As formulated in Equation 13, the total loss of a mask R-CNN model is defined as the sum of three individual losses, namely, the classification loss, the bounding box regression loss, and the mask loss (Zhao et al., 2019).

$$L = L(\text{classification}) + L(\text{bounding box regression}) + L(\text{mask}) \quad (13)$$

Equation 14 denotes the classification loss. In this Equation, p_i is the predicted probability of anchor i being an object, and p_i^* is the ground truth anchor label of an object.

The regressor is configured to learn scale-invariant transformation between two centers and log-scale transformation between widths and heights of the ground truth bounding boxes and predicted bounding boxes. This transformation is illustrated in Figure 38. The ground truth box coordinates are $g = (g_x, g_y, g_w, g_h)$ where x and y are the coordinates of the center, w is width, and h is height of the box, and predicted bounding box coordinates are $p = (p_x, p_y, p_w, p_h)$. Equations 15 to 18 denote the transformations, all of which taking p as the input. Finally, the bounding box regression loss function can be written as Equation 19.

$$L_{cls}(P_i, P_i^*) = -P_i^* \log P_i - (1 - P_i^*) \log (1 - P_i) \quad (14)$$

$$\hat{g}_x = p_h d_x(p) + p_x \quad (15)$$

$$\hat{g}_y = p_h d_y(p) + p_y \quad (16)$$

$$\hat{g}_w = p_w \exp(d_w(p)) \quad (17)$$

$$\hat{g}_h = p_h \exp(d_h(p)) \quad (18)$$

$$L(\text{bounding box regression}) = \sum_{i \in \{x, y, w, h\}} (t_i - d_i(p))^2 + \lambda \|w\|^2 \quad (19)$$

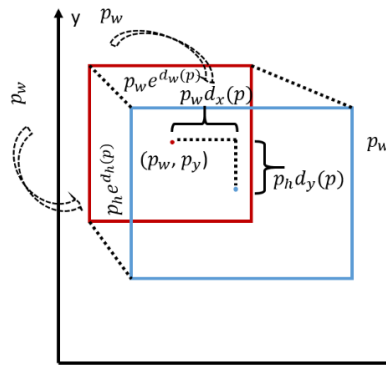


Figure 38 Transformation between Ground Truth and Prediction Bounding Boxes.

The mask loss is defined as the average binary cross-entropy loss, as expressed by Equation 20. In this Equation, y_{ij} represents a cell (i, j) that belongs to the ground truth mask for the region of size $m*m$. It uses the classification branch for category prediction and only relies on the ground truth class. \hat{y}_{ij}^k is the predicted value of the same cell in the mask learned for the ground truth class k .

$$L(mask) = -\frac{1}{m^2} \sum_{1 \leq i, j \leq m} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^k)] \quad (20)$$

Figures 39 illustrates the total loss, classification loss, bounding box regression loss, and mask loss per epoch for the trained Mask R-CNN model.

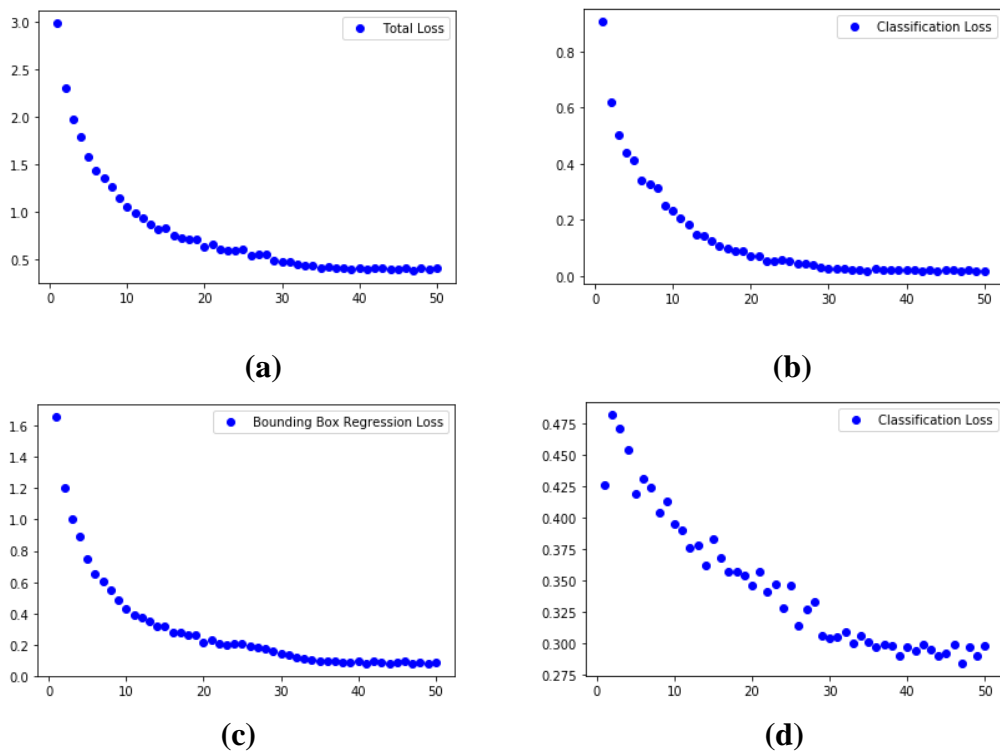


Figure 39 Loss Function per Epoch (a) Total (b) Classification (c) Bounding Box Regression (d) Mask.

The testing set contains 95 images of oil spill, taken from three different altitudes, namely high altitude (i.e., satellite imagery), low altitude (i.e., drone imagery), and first person. Figure 40 presents the distribution of the number of images based on the altitude.

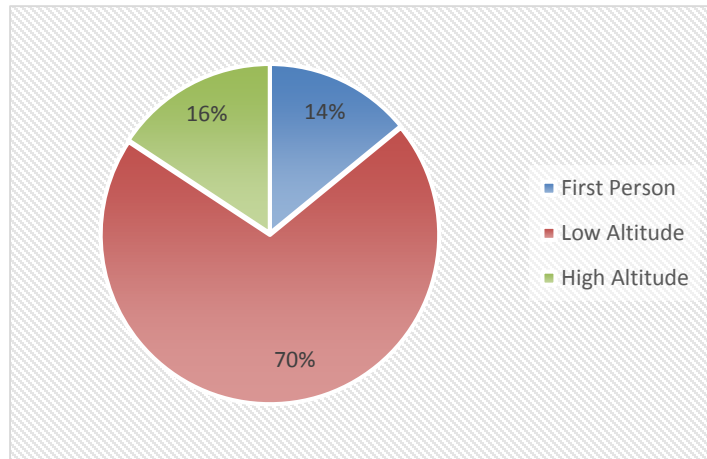


Figure 40 Distribution of the Number of Oil Spill Images Based on Altitude.

As before, hyperparameters used for both training and testing are chosen empirically, and listed in Table 8. Figure 41 presents sample ground truth images from the test dataset, and their corresponding detections marked with bounding boxes and masks. The mean average precision is 15.64% using transfer learning, and 12.56% when training from scratch.

Table 8 Hyperparameters Used for the Mask R-CNN Model.

Hyperparameter	Value
Batch size	1
Epochs	50
Steps	300
Score threshold	0.05
Binarize Threshold	0.5
IoU threshold	0.5

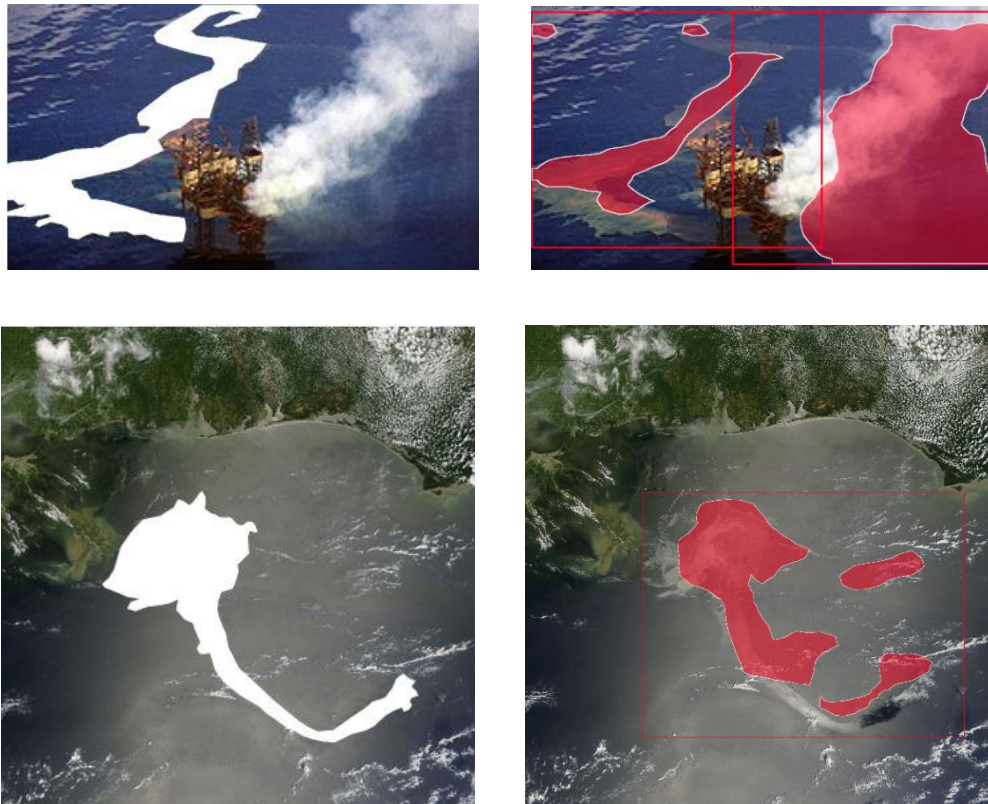


Figure 41 Sample Images from the Test Dataset (Left) and Corresponding Detected Images by Mask R-CNN (Right).
 (Images reprinted with permission from: [Montara Oil Spill- August 25, 2009](#) by Flickr user *SkyTruth*, CC-BY-NC-SA-2.0.; [Gulf of Mexico Oil Spill](#) by *Rob Gutro*.)

Object Segmentation Performance

To evaluate the performance of this model, each image from the testing set is assessed using pixel-level evaluation metrics, including precision, recall, IoU, F1, and Dice coefficient. Most of these metrics were defined in the previous Chapter. The definition of IoU for the pixel-level evaluation is given below:

Intersection over union (IoU), a widely used metric for instance segmentation problems, is defined as the ground truth and prediction intersection divided by their union and is

calculated using Equation 21 (Alipour et al., 2019). Figure 42 illustrates the positive and negative classes for pixel-level evaluation.

$$IoU = \frac{TP}{TP+FP+FN} \quad (21)$$

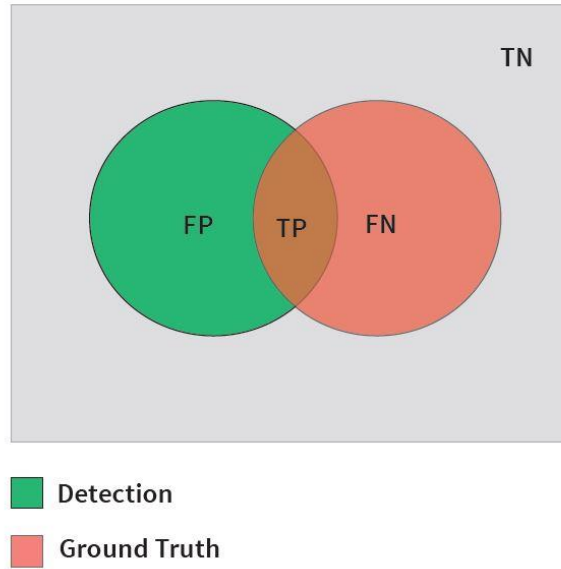


Figure 42 Positive and Negative Classes Based on the Detection and the Ground Truth.

Table 9 presents the pixel-level evaluation results based on different performance metrics, such as precision, recall, and IoU. According to this Table, the average precision for all the images is 62%. In other words, on average, for an image belonging to the testing set, the ratio of the pixels that are detected correctly (i.e., pixels that contained oil spill and are detected as oil spill) to the total positive classes (i.e., the sum of pixels that contained an oil spill and are detected as an oil spill and the pixels that do not contain any oil spill but are detected as an oil spill) is 62%.

Table 9 Performance Metrics Used for the Pixel-Level Evaluation on the Mask R-CNN Model.

	Precision	Recall	IoU	F1	Dice Coeff.
1	0.803	0.877	0.722	0.838	0.838
2	0.803	0.596	0.589	0.684	0.742
3	0.339	0.416	0.230	0.373	0.373
4	0.799	0.335	0.309	0.472	0.472
5	0.848	0.765	0.673	0.805	0.805
6	0.742	0.796	0.624	0.768	0.768
7	0.540	0.765	0.463	0.633	0.633
8	0.914	0.889	0.820	0.901	0.901
9	0.987	0.936	0.925	0.961	0.961
10	0.616	0.970	0.605	0.754	0.754
.
.
.
85	0.847	0.754	0.664	0.798	0.798
86	0.379	0.921	0.367	0.537	0.537
87	0.254	0.872	0.244	0.393	0.393
88	0.002	0.103	0.002	0.005	0.005
89	0.493	0.184	0.154	0.268	0.268
90	0.498	0.520	0.341	0.509	0.509
91	0.205	0.728	0.190	0.320	0.320
92	0.901	0.907	0.826	0.904	0.904
93	0.921	0.882	0.821	0.901	0.901
94	0.170	0.446	0.140	0.246	0.246
95	0.945	0.893	0.849	0.918	0.918
Average	62%	71%	49%	60%	60%

Similarly, the average recall for all the images is 71%. In other words, on average, for an image belonging to the testing set, the ratio of the pixels that are detected correctly (i.e., pixels that contain oil spill and are detected as oil spill) to the total number of TPs and FNs (i.e., the sum of pixels that contain an oil spill and are detected as an oil spill and

the pixels that contain an oil spill but are detected as no spill) is 71%. In addition, the average IoU is 49%, which indicates that on average, the ratio of TPs to the sum of TPs, FPs, and FNs in an image is 49%. Lastly, the average F1 score and Dice coefficient is 60%.

Figure 43 summarizes the distribution of the precision, recall, IoU, F1 Score, and Dice coefficient for the testing set images in pixel level. These metrics are used to evaluate the pixel-level performance of the Mask R-CNN model for each image belonging to the testing set. Evaluation results for each image using these metrics were previously presented in Table 9.

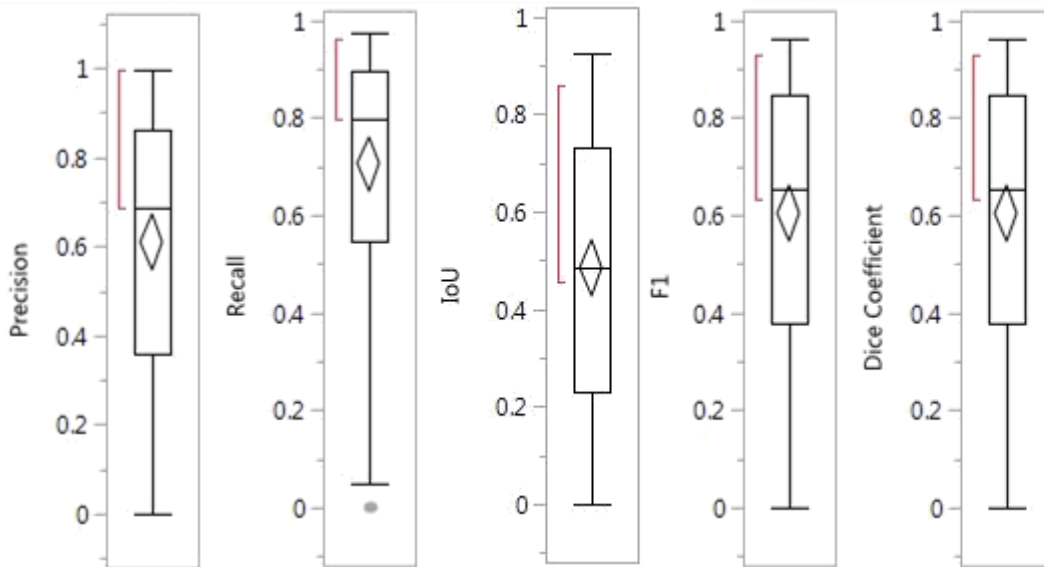


Figure 43 Box Plots Summarizing Performance Metrics of Pixel-Level Segmentation of Test Images.

According to this Figure, the maximum precision among all the testing set images is 0.99, and the minimum precision is 0. Also, one-fourth of the images have a precision

lower than 0.35, and half of the images have a precision lower than 0.68. The precision of three-fourth of the images is less than 0.86, and the mean precision is 0.61. The standard deviation (SD) of the precision values is 0.30. Similarly, the maximum recall among all the testing set images is 0.97, and the minimum recall is 0. Also, one-fourth of the images have a recall lower than 0.54, and half of the images have a recall lower than 0.79. The recall of three-fourth of the images is less than 0.89, and the mean recall is 0.70. The SD of the recall values is 0.26.

It can also be observed that the maximum IoU among all the testing set images is 0.92, and the minimum IoU is 0. Also, one-fourth of the images have an IoU lower than 0.22, and half of the images have an IoU lower than 0.48. The IoU of three-fourth of the images is less than 0.73, and the mean IoU is 0.48. The SD of the IoU values is 0.28. Moreover, the maximum F1 score among all the testing set images is 0.96, and the minimum F1 score is 0. Also, one-fourth of the images have a F1 score lower than 0.37, and half of the images have a F1 score lower than 0.65. The F1 score of three-fourth of the images is less than 0.84, and the mean F1 score is 0.60. The SD of the F1 scores is 0.27. Finally, the maximum Dice coefficient among all testing set images is 0.96, and the minimum Dice coefficient is 0. Also, one-fourth of the images have a Dice coefficient lower than 0.37, and half of the images have a Dice coefficient lower than 0.65. The Dice coefficient of three-fourth of the images is less than 0.84, and the mean Dice coefficient is 0.60. The SD of the Dice coefficient values is 0.30. Table 10 summarizes these results.

Table 10 Measures of Central Tendency and Variability for Precision and Recall of Testing Set Images in Pixel Level.

	Max	Min	1st Quartile	Median	3rd Quartile	Mean	SD
Precision	0.99	0.00	0.35	0.68	0.86	0.61	0.30
Recall	0.97	0.00	0.54	0.79	0.89	0.70	0.26
IoU	0.92	0.00	0.22	0.48	0.73	0.48	0.28
F1	0.96	0.00	0.37	0.65	0.84	0.60	0.27
Dice Coeff.	0.96	0.00	0.37	0.65	0.84	0.60	0.27

To take a closer look at the evaluated images, three examples are presented in Figure 44. Figure 44(a) corresponds to the ground truth image number 48 in the testing set, and Figure 44(b) presents the corresponding detection by the mask R-CNN model. According to the findings, precision, recall, IoU, F1 score, and Dice coefficient for this image are 0.503, 0.911, 0.479, 0.648, and 0.648, respectively. Figure 44(c) corresponds to the ground truth image number 38 in the testing set, and Figure 44(d) presents the corresponding detection by the mask R-CNN model. According to the findings, precision, recall, IoU, F1 score, and Dice coefficient for this image are 0.795, 0.808, 0.669, 0.801, and 0.801, respectively. Lastly, Figure 44(e) corresponds to the ground truth image number 94 in the testing set, and Figure 44(f) presents the corresponding detection by the mask R-CNN model. According to Table 9, precision, recall, IoU, F1 score, and Dice coefficient for this image are 0.170, 0.446, 0.140, 0.246, and 0.246, respectively.

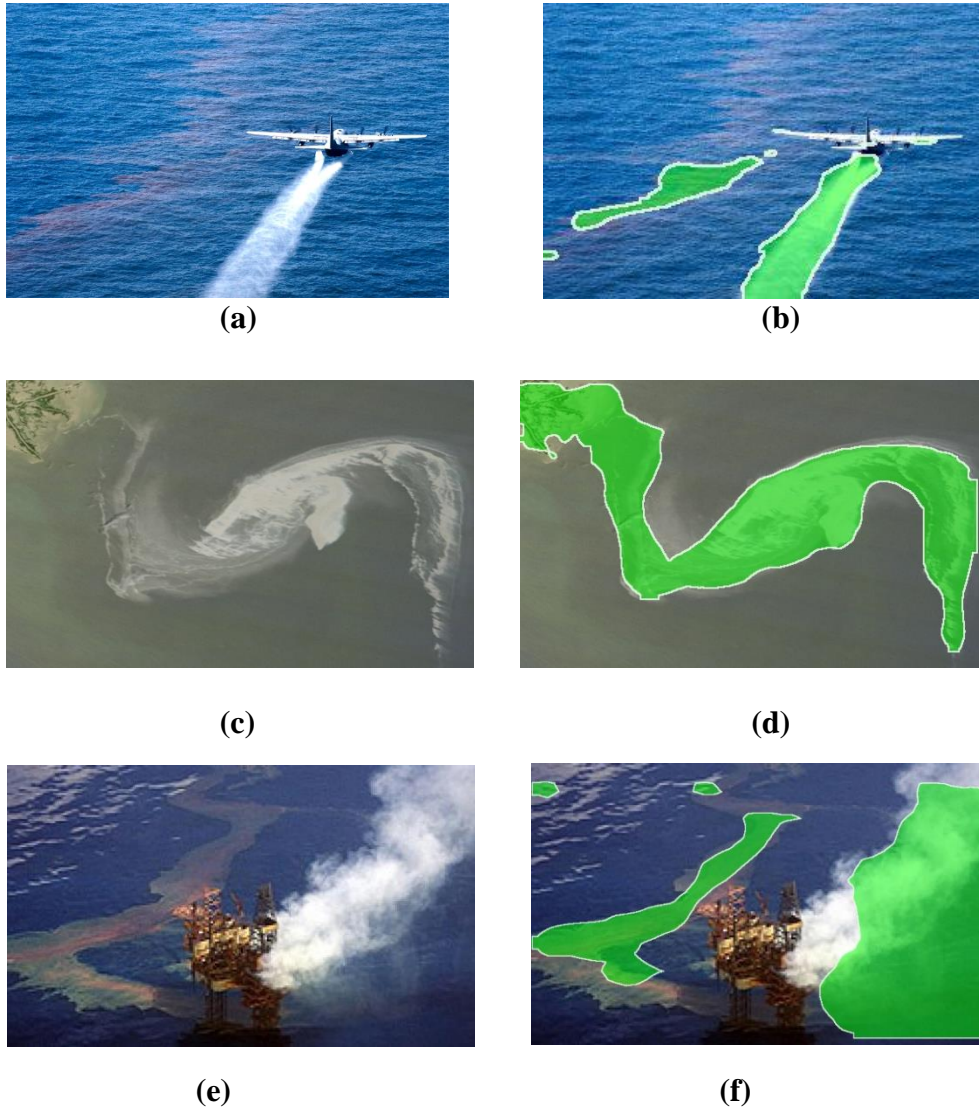


Figure 44 Sample Images from the Testing Set (Left) and Their Corresponding Detection Images by Mask R-CNN (Right).

(Images reprinted with permission from: [Fighting the Oil Slick](#) by *Adrian Cadiz*; [Gulf Oil Spill Creeps Towards Mississippi Delta Detail](#) 2010-04-29 by *Jesse Allen*; [Montara Oil Spill- August 25, 2009](#) by Flickr user *SkyTruth*, CC-BY-NC-SA-2.0.)

As mentioned earlier, the dataset consists of images from three different altitudes, including high altitude (i.e., satellite imagery), low altitude (i.e., drone imagery), and first-person images (i.e., images that are taken from the perspective of a person). To further evaluate the performance of the mask R-CNN model, the trained model (which is trained on images from all these altitudes) is tested separately on images from each group (satellite imagery, drone imagery, and first person). The model is then evaluated, using the evaluation metrics that were previously explained, such as precision, recall, and IoU. Table 11 summarizes the evaluation results. The first row presents the model performance in image level, using mean average precision. The following rows present the model performance based on pixel-level evaluation. Figure 45 draws a comparison between pixel-level model evaluation results (based on various metrics), based on the altitude of the testing set images.

Table 11 Performance Metrics of the Mask R-CNN Model Tested on Images from Different Altitudes.

	Satellite	Drone	First Person
mAP (Image level)	%9.97	%30.72	%47.50
Precision	0.506	0.573	0.545
Recall	0.546	0.686	0.689
Intersection over Union (IoU)	0.335	0.457	0.466
F1	0.525	0.592	0.582
Dice Coefficient	0.589	0.577	0.582

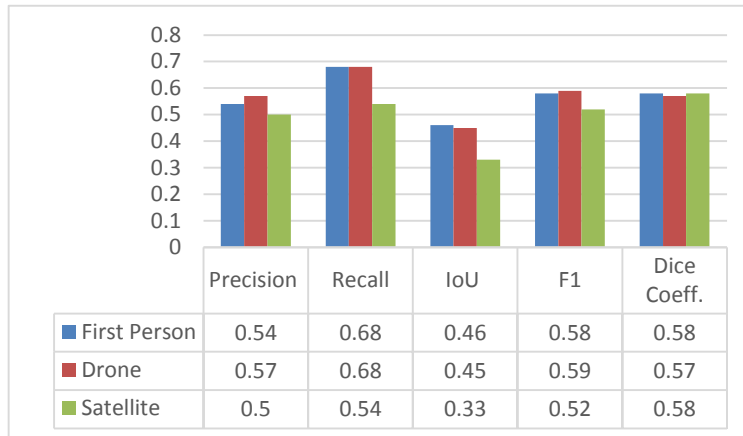


Figure 45 Pixel-Level Performance Metrics of the Mask R-CNN Model Tested on Images from Different Altitudes.

It must be noted that for any given test image, the generated masks marking the boundaries of adjacent oil spills may be separated from each other, as the model predicts each mask independent from other predictions in the neighboring areas. A pre-processing step is thus created to combine these fragmented masks prior to measuring the performance of the model in oil spill segmentation. Figure 46 shows three examples where separate detected masks, i.e., Figure 46(a) and (c) are merged to create a single mask, illustrated in Figures 46(b) and (d).

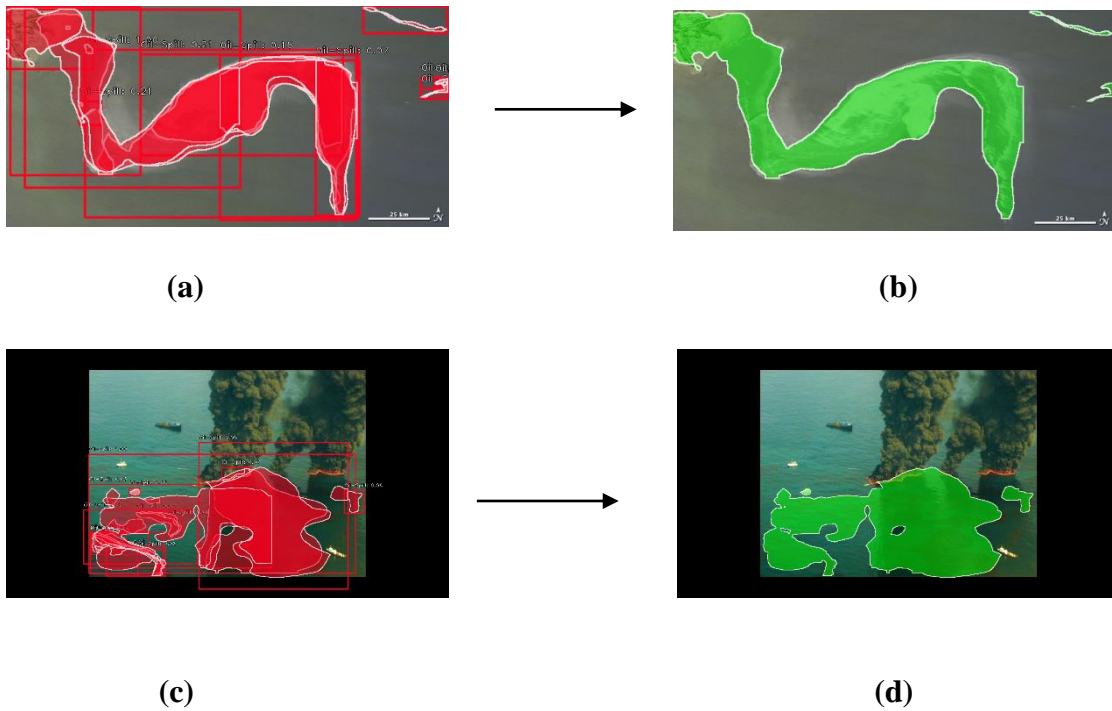


Figure 46 Preprocessing of Images from the Testing Set to Marge Separate Masks into One Single Mask.

(Images reprinted with permission from: [Gulf Oil Spill Creeps Towards Mississippi Delta Detail 2010-04-29](#) by *Jesse Allen*; [Controlled Burn of Oil on May 19th](#) by *John Kepsimelis*.)

CHAPTER V

CONCLUSIONS

Limitations

Although the mask R-CNN (Gkioxari et al., 2018) model deployed in this study is the most recent DL model (introduced in 2017) used for instance segmentation, and one of the best models for instance segmentation for complex objects (in this case, oil spills), it still cannot predict oil spills with high accuracy in certain cases. To start with, one of the issues that confuses the model and results in false positives is the presence of low RGB contrast between some oil spills and the background (i.e., water). Figure 47 shows an example of this issue where part of the background (i.e., water) is mistakenly detected as oil spill.



Figure 47 Example of False Positive Due to Low Contrast between Oil Spill and Background.

(Images reprinted with permission from: [Fighting the Oil Slick](#) by *Adrian Cadiz*.)

The other limitation of this study is the scarcity of the data obtained through web mining, since not all oil spill incidents receive wide news or social media coverage. Existing images on the public domain represent mostly some of the major oil spills (e.g.,

BP and Exxon Valdez) and those belong only to specific geographic locations. While some existing datasets from National Aeronautics and Space Administration (NASA) include larger numbers of oil spill images, they mostly contain satellite imagery. However, as previously explained, the key motivation behind this research is to use RGB drone imagery for model training and testing, since satellite imagery is expensive to attain and the frequency in which the data can be obtained is not enough.

More importantly, the concave and arbitrary shapes of oil spills make the instance segmentation more complicated for the model. Similar problems can be found in other domains where objects of interest have arbitrary shapes, such as medicine (to detect the boundaries of tumors) (Moeskops et al., 2016; Milletari et al., 2017; Roa et al., 2017; Farag et al., 2017; Brebisson and Montana, 2015) and morphology (Mancha et al., 2010). In particular, to this research, in certain cases, the reflection of objects on the water is mistaken for oil spills due to the darkness of the shades.

In addition, in some images, depending on the view angle and brightness, different shades of blue in the water can be mistaken for oil spills. As seen in Figure 48, some portions of the water are marked as oil spill by the model despite the fact that they do not contain any oil.



Figure 48 Example of False Positive Due to the Color of the Background.
(Images reprinted with permission from: Montara Oil Spill- August 25, 2009 by Flickr user *SkyTruth*, CC-BY-NC-SA-2.0.)

Future Work

Potential directions of future work in this research include adding mapping functionalities and benchmarking results against common industry practices. To this end, geographic information system (GIS) can be used to correspond detected oil spills and other objects of interest (e.g., vessels, cleaning crew) to geocoded locations on the map. Decision support system (DSS) can be also utilized to send this information to different stakeholders and authorities to take proper actions. As illustrated in Figure 49, following oil spill detection, the location of the oil spill will be mapped and the information will be communicated. Consequently, drones can be deployed to the oil spill location to retrieve more information and facilitate timely response to the accident.

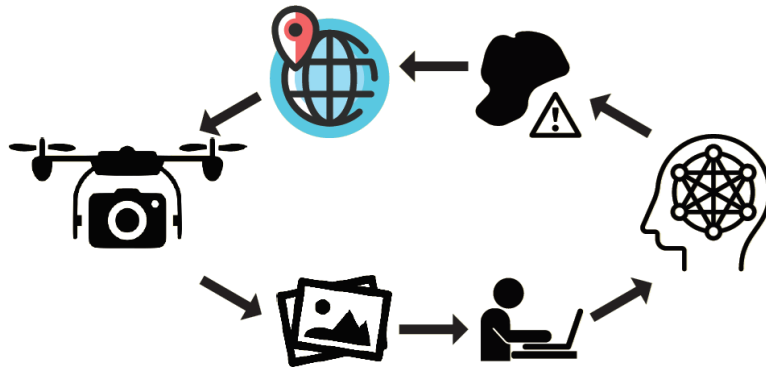


Figure 49 Schematic Diagram of the Future Direction of this Research.

Mapping can be done using different methods. For example, text mining can be used to extract location information from the metadata associated with a given image. Text mining is an area of research that tries to derive information such as patterns, trends, and semantic relationships embedded in textual data (Feldman and Sanger, 2007). Text mining has been implemented in different areas, including natural language processing (Gupta and Lehal, 2009), knowledge discovery (Mack and Henenberger, 2002), records management (Lee et al., 2007), patent mapping (Fattori et al., 2003), and social networking (Aggarwal, C. C. and Zhai, C., 2012). In this research, if the image includes metadata containing among others, coordinates information, that information can be retrieved through text mining and it can be pinned to the map.

In addition, triangulation can be used to map the location of an oil spill. Triangulation is the process of determining the location of an unknown point by forming triangles to it from known points (Hartley and Sturm, 1997). Figure 50 presents the schematic diagram of triangulation. In this method, the location of an oil spill can be

determined, using the coordinates of a known fixed object, such as an oil rig. However, the applicability of this method may be limited to cases where there are visible landmarks in the image. Figure 51 presents an example of a practical scenario (from the dataset developed in this Thesis) that is suitable for triangulation and mapping purposes. In this Figure, the location of oil spills could be calculated using the known location of the oil rig.

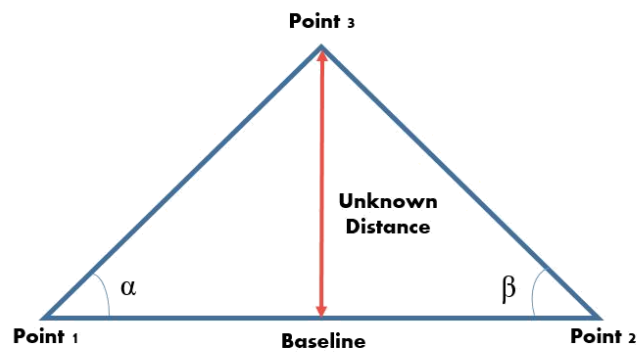


Figure 49 Schematic Diagram of Triangulation.

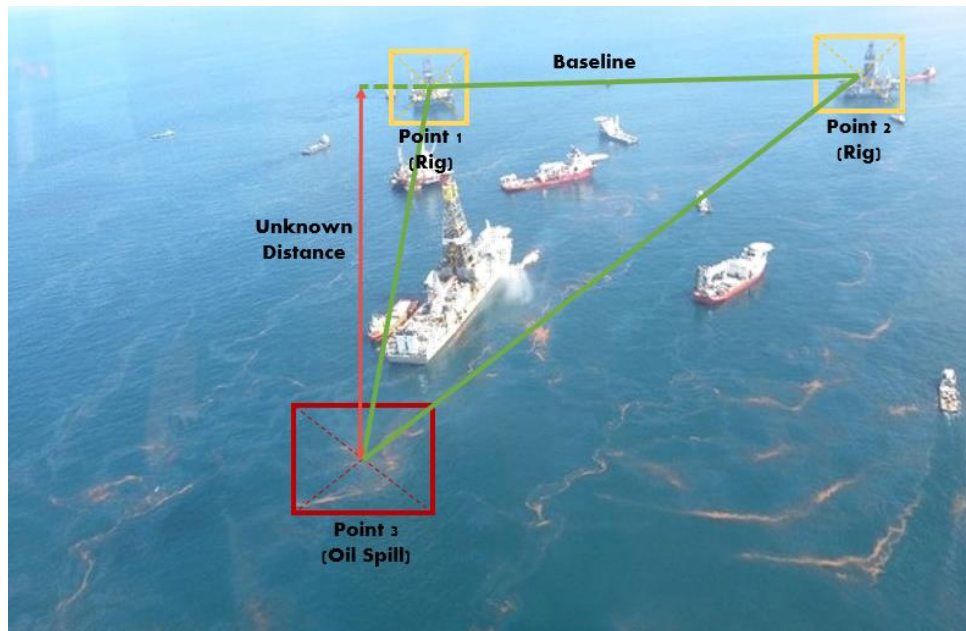


Figure 50 Example of a Practical Scenario (From the Dataset Developed in this Thesis) Suitable for Triangulation and Mapping Purposes.

(Images reprinted with permission from: [View of Deepwater Horizon Oil Spill in May 2020](#) by Flickr user *Office of Response and Restoration*, CC-BY-2.0.)

REFERENCES

- Abii, T.A. and Nwosu, P.C. (2009). "The effect of oil spillage on the soil of Eleme in Rivers State of the Niger-Delta area of Nigeria." *Res. J. Environ. Sci.*, 3(3), 316-320.
- Aggarwal, C.C. and Zhai, C. (2012). *Mining text data*. Springer Science & Business Media, New York, NY.
- Aguilera, F., Mendez, J., Pasaro, E., and Laffon, B. (2010). "Review on the effects of exposure to spilled oils on human health." *J. Appl. Toxicol.*, 30(1), 291-301.
- Akib, A.M., Saad, N., and Asirvadam, V. (2011). "Pressure point analysis for early detection system." *CSPA*, Penang, Malaysia, 103-107.
- Alpaydin, E. (2010). *Introduction to machine learning*, 2nd edition, MIT Press, London, England.
- Arnab, A. and Torr, P.H.S. (2017). "Pixel wise instance segmentation with a dynamically instantiated network." *CVPR*, Honolulu, HI, 441-450.
- Avelino, A.M., de Paiva, J.A., Da Silva, R.E.F., and Araujo, G.J.M. (2009). "Real-time leak detection system applied to oil pipelines using sonic technology and neural networks." *35th Annu. Conf. IEEE*, Porto, Portugal, 2109-2114.
- Awad, M. and Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Apress, London, England.
- Bengio, Y. (2009). "Learning deep architectures for AI." *Found. Trends Mach. Learn.*, 2(1), 1-127.

- Bengio, Y. (2012). "Deep learning of representations for unsupervised and transfer learning." *JMLR*, 27, 17-36.
- Bengio, Y., Courville, A., and Vincent, P. (2013). "Representation learning: a review and new perspectives." *IEEE T. Pattern Anal.*, 35(8), 1798-1828.
- Berry, M.W. (2004). "Survey of text mining." *Computing Reviews*, 45(9), pp. 548-560.
- Beyer, J., Trannum, H.C., Bakke, T., Hodson, P.V., and Collier, T.K. (2016). "Environmental effects of the Deepwater Horizon oil spill: a review." *Mar. Pollut. Bull.*, 110, 28-51.
- Bishop, C.M. (2006). *Pattern recognition and machine learning*, 1st edition. Springer, Berlin-Heidelberg, Germany.
- Blesito, S., Lombardi, P., Andreussi, P., and Banerjee, S. (1998). "Leak detection in liquefied gas pipelines by artificial neural networks." *AICHE J.*, 2675-2688.
- Bloch, H., Rafiq, S., and Salim, R. (2015). "Economic growth with coal, oil and renewable energy consumption in China: prospects for fuel substitution." *Econ. Model.*, 44, 104-115.
- Brekke, C. and Solberg, A.H.S. (2005). "Oil spill detection by satellite remote sensing." *Remote Sens. Environ.*, 95, 1-13.
- Cai, Z. and Vasconcelos, N. (2018). "Cascade R-CNN: delving into high quality object detection." *CVPR*, Salt Lake City, Utah, 6154-6162.
- Campbell, D., Cox, D., Crum, J., Foster, K., Christie, P., and Brewster, D. (1993). "Initial effects of the grounding of the tanker Braer on health in Shetland." *Br. Med. J.*, 307, 1251-1255.

- Cheng, J. and Baldi, P. (2006). "A machine learning information retrieval approach to protein fold recognition." *Bioinformatics*, 22(12), 1456-1463.
- Chollet, F. (2018). *Deep learning with python*, 1st edition. Manning Publications, Shelter Island, NY.
- Claesen, M. and Moor, B.D. (2015). "Hyperparameter search in machine learning." *arXiv:1502.02127*.
- Cococcoioni, M., Corucci, L., Masini, A., and Nardelli, F. (2012). "SVME: an ensemble of support vector machines for detecting oil spills from full resolution MODIS images." *Ocean Dynam.*, 62, 449-467.
- Cohen, M.J. (1995). "Technological disasters and natural resource damage assessment: and evaluation of the Exxon Valdez oil spill." *Land Econ.*, 71(1), 65-82.
- Covas, D. and Ramos, H. (1999). "Practical methods for leakage control, detection, and location in pressurized systems." *BHR group conference series publication*, 37, 135-152.
- Covas, D., Ramos, H., and De Almeida, A.B. (2005). "Standing wave difference method for leak detection in pipeline systems." *J. Hydraul. Eng.*, 131(12), 1106-1116.
- Dai, W., Yang, Q., Xue, G.R., and Yu, Y. (2007). "Boosting for transfer learning." *ICML*, Corvallis, OR.
- Das, S., Dey, A., Pal, A., and Roy, N. (2015). "Applications of artificial intelligence in machine learning: review and prospect." *Int. J. Comput. Appl.*, 115(9), 31-41.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L. (2009). "ImageNet: a large-scale hierarchical image database." *CVPR*, Miami, FL, 248-255.

- Doerffer, J. W. (2013). *Oil spill response in the marine environment*. Elsevier, Tarrytown, NY.
- Dubansky, B., Whitehead, A., Miller, J.T., Rice, C.D., and Galvez, F. (2013). “Multitissue molecular, genomic, and developmental effects of the Deepwater Horizon oil spill on resident Gulf Killifish (*Fundulus Grandis*).” *Environ. Sci. Technol.*, 47, 5074-5082.
- Dunham, M.H. (2006). *Data mining: introductory and advanced concepts*. Pearson Education, Upper Saddle River, NJ.
- Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J., and Zisserman, A. (2010). “The PASCAL visual object classes (VOC) challenge.” *Int. J. Comput. Vis.*, 88(2), 303-338.
- Fattori, M., Pedrazzi, G., and Turra, R. (2003). “Text mining applied to patent mapping: a practical business case.” *World Pat. Inf.*, 25, 335-342.
- Feldman, R. and Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, New York, NY.
- Ferraro, G., Bernardini, A., David, M., Meyer-Roux, S., Mullenhoff, O., Perkovic, M., Tarchi, D., and Topouzelis, K. (2007). “Towards an operational use of space imagery for oil pollution monitoring in the Mediterranean basin: a demonstration in the Adriatic Sea.” *Mar. Pollut. Bull.*, 54, 403-422.
- Ferraz, I.N., Garcia, A.C.B., and Bernardini, F.C. (2009). “Artificial neural networks ensemble used for pipeline leak detection systems.” *IPC*, Calgary, Alberta, Canada, 739-747.
- Fingas, M. (2011). *Oil spill science and technology: prevention, response, and cleanup*. Elsevier, Burlington, MA.

- Fingas, M. (2012). *The basics of oil spill cleanup*, 3rd edition. CRC Press, Boca Raton, FL.
- Fingas, M.F. and Brown C.E. (1997). "Review of oil spill remote sensing." *Spill Sci. Technol. Bull.*, 4(4), 199-208.
- Fingas, M. and Brown, C.E. (2017). "A review of oil spill remote sensing." *Sensors*, 18(91), 1-18.
- Fiscella, B., Giancaspro, A., Nirchio, F., Pavese, P., and Trivero, P. (2000). "Oil spill detection using marine SAR images." *Int. J. Remote Sens.*, 21(18), 3561-3566.
- Fuchs, H.V. (1991). "One year of experience leak detection by acoustic signal analysis." *Appl. Acoust.*, 3(1), 1-19.
- Garza-Gil, M.D., Prada-Blanco, A., and Vazquez-Rodriguez, M.X. (2006). "Estimating the short-term economic damages from the Prestige oil spill in the Galician fisheries and tourism." *Ecol. Econom.*, 58(1), 842-849.
- Gately, D., Al-Yousef, N., and Al-Sheikh, H.M.H. (2012). "The rapid growth of domestic oil consumption in Saudi Arabia and the opportunity cost of oil exports foregone." *Energy Policy*. 47(1), 57-68.
- Geiger, G. (2008). *Principles of leak detection*, 1st edition. Krohne Oil & Gas, Breda, the Netherlands.
- Geiger, G. and Werner, T. (2003). "Leak detection and location – a survey." *Proceedings of Pipeline Simulation Interest Group (PSIG) Annual Meetings*, Bern, Switzerland, 1-11.
- Geiger, G., Vogt D., and Tetzner, R. (2006). "State-of-the-art leak detection and localization." *Oil Gas Eur. Mag.*, 32(4), 193-210.

- Ghazali, M.F. (2012). "Leak detection using instantaneous frequency analysis." PhD Dissertation, University of Sheffield, UK.
- Girard-Ardhuin, F., Mercier, G., and Garello, R. (2003). "Oil slick detection by SAR imagery: potential and limitation." *Oceans* 5(1), 164-169.
- Girshick, R. (2015). "Fast R-CNN." *ICCV*, Las Condes, Chile, 1440-1448.
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation." *CVPR*, Columbus, OH, 580-587.
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2016). "Region-based convolutional networks for accurate object detection and segmentation." *IEEE T. Pattern Anal.*, 38(1), 142-158.
- Goldstein, B.D., Osofsky, H.J., and Lichtveld, M.Y. (2011). "The Gulf oil spill." *N. Engl. J. Med.* 364(14), 1334-1348.
- Gupta, V. and Lehal, G.S. (2009). "A Survey of text mining techniques and applications." *J. Emerg. Technol. Web Intell.* 1(1), 60-76.
- Harrington, P. (2012). *Machine learning in action*. Manning Publication, Shelter Island, NY.
- Hartley, R.I. and Sturm, P. (1997). "Triangulation." *Comput. Vis. Image Und.* 68(2), 146-157.
- He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2018). "Mask R-CNN." *CVPR*, Salt Lake City, UT, 2961-2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE T. Pattern Anal.*, 37(9), 1904-1916.

- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition." *CVPR*, Las Vegas, NV, 770-778.
- He, K. (2009). "Neuro based classification of gas leakage sounds in pipeline." *Proc. Int. Conf. Networking Sensing Control*, Okayama, Japan, 298-302.
- Hinton, G.E. and Salakhutdinov, R.R. (2006). "Reducing the dimensionality of data with neural networks." *Science*, 313(5786), 504-507.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.R., Jaitly, N., Senior, A., Vanhouke, V., Nguyen, P., Kingsbury, B., and Sainath, T. (2012). "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups." *IEEE Signal Process. Mag.*, 29(6), 82-97.
- Hou, Z., Hu, Q., and Nowinski, W.L. (2006). "On minimum variance thresholding." *Pattern Recognit. Lett.*, 1732-1743.
- Huang, M., Yu, W., and Zhu, D. (2012). "An improved image segmentation algorithm based on the Otsu method." *13th ACIS Int. Conf. Software Eng. Artificial Intelligence, Networking Parallel/Distributed Comput.*, Kyoto, Japan, 135-139.
- Indiveri, G. and Liu, S.C. (2015). "Memory and information processing in neuromorphic systems." *Proc. IEEE*, 103(8), 1379-1397.
- Janjua, N.Z., Kasai P.M., Nawaz, H., Forooqui, S.Z., Khuwaja, U.B., Hassan, N., Jafri, S. N., Lutfi, S.A., Kadir, M.M., and Sathiakumar, N. (2006). "Acute health effects of the Tasman Spirit oil spill on residents of Karachi, Pakistan." *BMC Pub Health*, 6(84).

- Jha, M.N., Levy, J., and Gao, Y. (2008). "Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance." *Sensors*,8(1), 236-255.
- Jiao, Z, Jia, G, and Cai, Y. (2019). "A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles." *Comput. Ind. Eng.*, 135(1), 1300-1311.
- Jin, H., Zhang, L., Liang, W., and Ding, Q. (2014). "Integrated leakage detection and localization model for gas pipelines based on the acoustic wave method." *J. Loss Prev. Process Ind.*, 27(1), 74-88.
- Jing, Y., Bian, Y., Hu, Z., Wang, L., and Xie, X.S. (2018). "Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era." *AAPS J.*, 20(58), 1-10.
- Kadhim, M. and Parry J.M. (1984). "The detection of mutagenic chemicals in the tissues of shellfish exposed to oil pollution." *Mutat. Res.*, 136(2), 93-105.
- Kapelan, Z.S., Savic, D.A., and Walters, G.A. (2003). "A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks." *J. Hydraul. Res.*, 41(5), 481-492.
- Kavukcuoglu, K., Fergus, R., and LeCun Y. (2009). "Learning invariant features through topographic filter maps." *CVPR*, Miami, FL, 1605-1612.
- Kingston, P.F. (2002). "Long-term environmental impact of oil spills." *Spill Sci. Technol. Bull.*, 7(1), 53-61.

- Kirscht, M. and Rinke, C. (1998). "3D reconstruction of buildings and vegetation from synthetic aperture radar (SAR) images." *IAP Workshop on Mach. Vision Appl.*, Chiba, Japan, 17-19.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). "Imagenet classification with deep convolutional neural networks." *NIPS*, 1097-1105.
- Kubat, M., Holte. R.C., and Matwin, S. (1998). "Machine learning for the detection of oil spills in satellite radar images." *Mach. Learn.*, 30, 195-215.
- Laur, H., Bally, P., Meadows, P., Sanchez, J., Schaettler, B., and Lopinto, E. (2002). "Derivation of the backscattering coefficient in ESA ERS SAR PRI products." *European Space Agency (ESA)*, Noordwijk, The Netherlands, ESA Document No.: ES-TN-RS-PMHL09(2).
- LeCun, Y., Bengio, Y., and Hinton, G.E. (2015). "Deep learning." *Nature*, 521(7553), 436-444.
- Lee, C.H., Kang, Y.A., Chang, K.J., Kim, C.H., Hur, J.I., Kim, J.Y., and Lee, J.K. (2010). "Acute health effects of the Hebei oil spill on the residents of Taean, Korea." *J. Prev. Med. Public Health*, 43(2), 166–173.
- Lee, J.Y., Moon, J.Y., and Kim, H.J. (2007). "Examining the intellectual structure of records management & archival science in Korea with text mining." *J. Korean Soci. Library Inf. Sci.*, 41(1), 345-372.
- Lee, S.U. and Chung, S.Y. (1990). "A comparative performance study of several global thresholding techniques for segmentation." *Comput. Gr. Image Process.*, 52, 171-190.

- Levy, B.S. and Nassetta, W.J. (2011). "The adverse health effects of oil spills: a review of the literature and a framework for medically evaluating exposed individuals." *Int. J. Occup. Environ. Health*, 17(2), 161-168.
- Li, Y., Chen, X., and Yang, B. (2002). "Research on web mining-based intelligent search engine." *Proc. Int. Con. Mach. Learn. Cyb.*, Beijing, China, 386-390.
- Liao, P., Chen, T., and Chung, P. (2001). "A fast algorithm for multilevel thresholding." *J. Inf. Sci. Eng.*, 17, 713-727.
- Lin, K.C. (2001). "Fast thresholding computation by searching for zero derivatives of image between-class variance." *27th Annu. Conf. IEEE Ind. Electron. Soci.*, Denver, CO, 393-397.
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., and Zitnick, L. (2014). "Microsoft COCO: common objects in context." *Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 740-755.
- Lin, T., Dollar, p., Girshick, R.B., He, K., Hariharan, B., and Belongie, S.J. (2017). "Feature pyramid networks for object detection." *CVPR*, Honolulu, HI, 2117-2125.
- Liu, D. and Yu, J. (2009). "Otsu method and k-means." *9th Int. Conf. Hybrid Intelligent Systems*, Shenyang, China, 344-349.
- Liu, P., Zhao, C., Li, X., He, M., and Pichel, W. (2010). "Identification of ocean oil spills in SAR imagery based on fuzzy logic algorithm." *Int. J. Remote Sens.*, 31(17-18), 4819-4833.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F.E. (2017). "A survey of deep neural network architectures and their applications." *Neurocomputing.*, 234, 11-26.

- Liu, X. and Wirtz, K.W. (2006). "Total oil spill costs and compensations." *Marit. Pol. Manag.*, 33(1), 49-60.
- Loureiro, M.L., Ribas, A., Lopez, E., and Ojea, E. (2006). "Estimated costs and admissible claims linked to the Prestige oil spill." *Ecol. Econom.*, 59(1), 48-63.
- Lyons, R.A., Temple, J.M., Evans, D., Fone, D.L., and Palmer, S.R. (1999). "Acute health effects of the Sea Empress oil spill." *Epidemiol Community Health Journal*, 53, 306-310.
- Mack, R. and Henenberger, M. (2002). "Text-based knowledge discovery: search and mining of life-sciences documents." *Drug Discov. Today*, 7(11), 89-98.
- Marques, M.M., Lobo, V., Batista, R., Almeida, J., Nunes, M. F., Ribeiro, R., and Bernardino, A. (2016). "Oil spills detection challenges addressed in the scope of the SEAGULL project." *IEEE Conf.*, Monterey, CA, 1-6.
- Meo S. A., Al-Drees A. M., Rasheed S, Meo, I. M., Al-Saadi, M. M., Ghani, H. A., and Alkandari, J.R. (2009). "Health complaints among subjects involved in oil cleanup operations during oil spillage from a Greek tanker, Tasman Spirit." *Int. J. Occup. Med. Env.*, 22(2), 143-148.
- Mera, D., Cotos, J.M., Varela-Pet, J., and Garcia-Pineda, O. (2012). "Adaptive thresholding algorithm based on SAR images and wind data to segment oil spills along the northwest coast of the Iberian Peninsula." *Mar. Pollut. Bull.*, 64, 2090-2096.
- Mera, D., Cotos, J.M., Varela-Pet, J., Rodriguez, P.G., and Caro, A. (2014). "Automatic decision support system based on SAR data for oil spill detection." *Comput. Geosci.*, 72, 184-191.

- Milletari, F., Ahmadi, S.A., Kroll, C., Plate, A., Rozanski, V., Maistre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Botzel, K., and Navab, N. (2017). "Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound." *Comput. Vis. Image Und.*, 164, 92-102.
- Min, S., Lee, B., and Yoon, S. (2017). "Deep learning in bioinformatics." *Brief. Bioinform.*, 18(5), 851-869.
- Mitchell, T. (1997). *Machine learning*. McGraw Hill Science, Engineering, and Math, Portland, OR.
- Morita, A., Kusaka, Y., Deguchi, Y., Moriuchi, A., Nakanaga, Y., Iki, M., Miyazaki, S., and Kawahara, K. (1999). "Acute health problems among the people engaged in the cleanup of the Nakhodka oil spill." *Environ. Res.*, 81(3), 185-194.
- Moss, L. (2010). "The 13 largest oil spills in history." Mother Nature Work Online Magazine, < <https://www.mnn.com/earth-matters/wilderness-resources/stories/the-13-largest-oil-spills-in-history> > (accessed 20 May 2019) .
- Murvay, P. and Silea, L. (2012). "A survey on gas leak prevention and localization techniques." *J. Loss Prev. Process Ind.*, 25(6), 966-973.
- Muttin, F. (2011). "Umbilical deployment modeling for tethered UAV detecting oil pollution from ship." *Appl. Ocean Res.*, 33(4), 332-343.
- Nath, N.D. and Behzadan, A.H. (2019). "Deep learning models for content-based retrieval of construction visual data." *J. Comput. Civil Eng.*, 66-73.

- Negro, M.C.G., Villasante, S., Penela, A.C., and Rodriguez, G.R. (2009). "Estimating the economic impact of the Prestige oil spill on the death coast (NW Spain) fisheries." *Mar. Policy.*, 33, 8-23.
- Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H.G., and Ogata, T. (2015). "Audio-visual speech recognition using deep learning." *Appl. Intell.*, 42(4), 722-737.
- Oquab, M., Bottou, L., Laptev, I. and Sivic, J. (2014). "Learning and transferring mid-level image representations using convolutional neural networks." *CVPR*, Columbus, OH, 1717-1724.
- Oquab, M., Bottou, L., Laptev, I. and Sivic, J. (2014). "Weakly supervised object recognition with convolutional neural networks." *NIPS*, 1-6.
- Otsu, N. (1979). "A threshold selection method from gray-level histograms." *IEEE Trans. Syst. Man. and Cybern.*, 9(1), 62-66.
- Palinkas, L.A. (2012). "A conceptual framework for understanding the mental health impacts of oil spills: lessons from the Exxon Valdez oil spill." *Psychiatry*, 75(3), 203-222.
- Palinkas, L.A., Petterson, J.S., Russell, J., and Downs, M.A. (1993). "Community patterns of psychiatric disorders after the Exxon Valdez oil spill." *Am. J. Psychiatry.*, 150(10), 1517-1523.
- Pan, S.J. and Yang, Q. (2010). "A survey on transfer learning." *IEEE T. Knowl. Data En.*, 22(10), 1345-1359.
- Pan, S.J., Kwok, J.T., and Yang, Q. (2008). "Transfer learning via dimensionality reduction." *Proc. 23rd AAAI Conf. AI*, Chicago, IL, 677-682.

- Powers, D.M. (2011). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." *J. Mach. Learn. Techn.*, 2(1), 37-63.
- Raimondi, V., Palombi, L., Lognoli, D., Masini, A., and Simeone, E. (2017). "Experimental tests and radiometric calculations for the feasibility of fluorescence lidar-based discrimination of oil spills from UAV." *Int. J. Appl. Earth Obs.*, 61, 46-54.
- Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A.Y. (2007). "Self-taught learning: transfer learning from unlabeled data." *ICML*, Corvallis, OR, 759-766.
- Redmon, J. and Farhadi, A. (2018). "YOLOv3: an incremental improvement." *arXiv: 1804.02767*.
- Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger." *CVPR*, Honolulu, HI, 7263-727.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection." *CVPR*, Las Vegas, NV, 779-788.
- Ritchie, B.W., Crotts, J.C., Zherer, A., and Volsky, G.T. (2013). "Understanding the effects of a tourism crisis: the impact of the BP oil spill on regional lodging demand." *J. Travel Res.*, 53(1), 12-25.
- Rodriguez-Trigo, G., Zock, J.P., and Montes, I.I. (2007). "Health effects of exposure to oil spills." *Arch. Bronconeumol. (English Edition)*, 43(11), 628-635.
- Romero, L.M., and Wilkelski, M. (2002). "Severe effects of low-level oil contamination on wildlife predicted by the corticosterone-stress response: preliminary data and a research agenda." *Spill Sci. Technol. Bull.*, 7(5-6), 309-313.

- Russell, S.J. and Norvig, P. (2010). *Artificial intelligence: a modern approach*, 3rd edition. Prentice Hall, Malaysia.
- Samuel, A. (1988). *Some studies in machine learning using the game of checkers*. Springer, New York, NY.
- Schmidhuber, J. (2015). "Deep learning in neural networks: an overview." *Neural Networks*, 61, 85-117.
- Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2013). "Overfeat: integrated recognition, localization and detection using convolutional networks." *arXiv:1312.6229*.
- Sezgin, M. and Sankur, B. (2004). "Survey over image thresholding techniques and quantitative performance evaluation." *J. Electron. Imaging.*, 13(1), 146-165.
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R.M. (2016). "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." *IEEE Trans. Med.*, 35(5), 1285-1298.
- Shu, Y., Li, J., Yousif, H., and Gomes, G. (2010). "Dark-spot detection from SAR intensity imagery with spatial density thresholding for oil spill monitoring." *Remote Sens. Environ.*, 114, 2026-2035.
- Sim, M.S., Jo, I.J., and Song, H.G. (2010). "Acute health problems related to the operation mounted to clean the Hebei Spirit oil spill in Taean, Korea." *Mar. Pollut. Bull.*, 60(1), 51–57.

- Simonyan, K. and Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition." *arXiv:1409.1556*.
- Solberg, A.H.S., Brekke, C., and Husoy, P.O. (2007). "Oil spill detection in RADARSAT and ENVISAT SAR images." *IEEE T. Geosci. Remote.*, 45(3), 746-755.
- Solberg, A.H.S., Storvik, G., Solberg, R., and Volden, E. (1999). "Automatic detection of oil spills in ERS SAR images." *IEEE T. Geosci. Remote.*, 37(4), 1916- 1924.
- Song, M., Hong, Y.C., Cheong, H.K., Ha, M., Kwon, H., Ha, E.H., Choi, Y., Jeong, W. C., Hur, J., Lee, S.M., and Kim, E.J. (2009). "Psychological health in residents participating in clean-up works of Hebei Spirit oil spill." *J. Prev. Med. Public Health*, 42(2), 82–88.
- Suarez, B., Lope, V., Gomez, B.P., Aragonés, N., Artalejo, F.R., Marques, F., Guzman, A., Vitoria, L.J., Carrasco, J.M., Moreno, J.M.M., Abente, G.L., and Pollan, M. (2005). "Acute health problems among subjects involved in the cleanup operation following the Prestige oil spill in Asturias and Cantabria (Spain)." *Environ. Res.*, 99(3), 413-424.
- Swedmark, M., Granmo, A., and Kollberg, S. (1973). "Effects of oil dispersants and oil emulsions on marine animals." *Water Res.*, 7, 1649-1972.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., (2015). "Going deeper with convolutions." *CVPR*, Boston, MA, 1-9.
- Valez, P., Johnsen, H.G., Steen, A., and Osikilo, Y. (2011). "Advancing oil spill preparedness and response techniques for arctic conditions." *Int. Oil Spill Conf. Proc.*, 11(1). American Petroleum Institute, abs105.

Ventikos, N.P. (2002). “Development of an evaluation model for the importance, the causes and the consequences of oil marine pollution: the case of maritime transport in the Greek seas and in the Gulf of Saronikos” National Technical University of Athens, Greece.

Ventikos, N. P., Vergetis, E., Psaraftis, H. N., and Triantafyllou, G. (2004) “A high-level synthesis of oil spill response equipment and countermeasures.” *J. Hazard. Mater.*, 107, 51-58.

Vespe, M., Ferraro, G., Posada, M., Greidanus, H., and Perkovic, M. (2011). “Oil spill detection using Cosmo-SkyMed over the Adriatic Sea: the operational potential.” *IGARSS 2011 Conf.*, Vancouver, Canada, 4403-4406.

Vitkovsky, J.P., Lambert, M.F., Simpson, A.R., and Liggett, J.A. (2007). “Experimental observation and analysis of inverse transients for pipeline leak detection.” *J. Water Res. Plan. Man.*, 133(6), 519-530.

Wadley, F.M. (1947). “Probit analysis: a statistical treatment of the sigmoid response curve.” *Annu. Entomol. Soc. Am.*, 67(4), 549–553.

Weiss, K., Khoshgoftaar, T. M., and Wang, D. D. (2016). “A survey of transfer learning.” *J. Big Data*, 3(9), 1-40.

West, J., Ventura, D., and Warnick, S. (2007). “Spring research presentation: a theoretical foundation for inductive transfer.” Brigham Young University, College of Physical and Mathematical Sciences.

Wu, G., Lu, W., Gao, G., Zhao, C., and Liu, J. (2016). “Regional deep learning model for visual tracking.” *Neurocomputing*, 175, 310-323.

- Xu, L., Li, J., and Brenning, A. (2014). "A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery." *Remote Sens. Environ.*, 141, 14-23.
- Yang, Z., Xiong Z., and Shao, M. (2010). "A new method of leak location for the natural gas pipeline based on wavelet analysis." *Energy J.*, 35(9), 3814- 3820.
- Yao, Y. and Doretto, G. (2010). "Boosting for transfer learning with multiple sources." *CVPR*, San Francisco, CA, 1855-1862.
- Yosinski, J. and Clune, J. (2014). "How transferable are features in deep neural networks?" *NIPS*, 3320-3328.
- Yu, D. and Deng, L. (2011). "Deep learning and its applications to signal and information processing" *IEEE Signal Process Mag.*, 28(1), 145-154.
- Zaart, A.E. and Ghosn, A.A. (2013). "SAR images thresholding for oil spill detection." *SIECPC*, Fira, Greece, 1-5.
- Zhang, J. (1996). "Designing a cost effective and reliable pipeline leak detection system." *Pipeline Rel. Conf.*, Houston, TX, 1-11.
- Zhang, J., Hoffman, A., Murphy, K., Lewis, J., and Twomey, M. (2013). "Review of pipeline leak detection technologies." *Proc. PSIG Annu.*, Prague, Czech Republic.
- Zhang, Y., Chen, S., Li, J., and Jin, S. (2014). "Leak detection monitoring system of long-distance oil spill pipeline based on dynamic pressure transmitter." *Measurement*, 49, 382-389.
- Zhao, Z.Q., Zheng, P., Xu, S., and Wu, X. (2019). "Object detection with deep learning: a review." *IEEE T. Neur. Net. Lear.*, 1-21.

Zhigang, C., Xiangjiao, L., and Zhihong, Y. (2010). "Leakage detection and localization method for pipelines in complicated conditions." *Proc. 29th Chinese Control Conf.*, Beijing, China, 4057- 4061.

Zou, G. and Chau, K.W. (2006). "Short- and long-run effects between oil consumption and economic growth in China." *Energy Policy*, 34(18), 3644-3655.