



## Deep learning-based oil spill detection with LWIR camera

Ho-Min Park<sup>1</sup> · Gyung-Soo Park<sup>2</sup> · Jung-Kn Kim<sup>3</sup> · Jae-Hoon Kim<sup>4</sup> · Seong-Dae Lee<sup>†</sup>

(Received November 30, 2021 ; Revised December 17, 2021 ; Accepted December 17, 2021)

**Abstract:** The ocean covers approximately 71% of the total surface area of Earth and plays a significant role in maintaining the environment and ecosystems. Oil spills are the largest source of pollution in the ocean, mainly Bunker C oil and diesel oil used as vessel fuels. Therefore, oil spill detection is essential for marine protection, which motivated this study. Detection with radar, which is based on electromagnetic waves, is achieved using satellite synthetic aperture radar (SAR); thus, real-time detection over a small range is difficult. Hence, in this study, an oil spill detection system based on thermal imaging using a long-wave infrared (LWIR) camera is proposed. The oil spill detection algorithm utilizes the You Only Look Once (YOLO) model, which is widely used for object detection. In addition, 1,644 thermal images were labeled to evaluate the proposed system. The training and test results showed an accuracy of 96.91% and false alarm rate of 8.33%. An improved detection performance can be expected from subsequent experiments using larger image datasets.

**Keywords:** Oil spill detection, YOLO, Object detection, Infrared imaging

### 1. Introduction

The ocean accounts for approximately 71% of the surface of Earth and plays a key role in maintaining various environments and ecosystems. A polluted ocean and its compromised ability to fulfil its roles will either exacerbate global warming or destroy marine ecosystems, triggering calamities to humanity. Therefore, the development of technologies for marine pollution detection is essential.

Oil is the largest source of pollution in the ocean and has severe impacts after discharge [1]. Bunker C and diesel oils, which are used as vessel fuels, are mainly spilled. Thus, the number of spills has increased along with the annual increase in vessel accidents [2].

Marine oil pollution detection technologies can be classified into two methods: electromagnetic wave-based detection using radar and image-based detection using cameras [3]. Image-based detection requires only a camera for recording and embedded equipment for image analysis [4]. In contrast, electromagnetic

wave-based detection requires the installation of equipment at the highest possible locations and the strong output power of an electromagnetic wave generator to increase the accuracy and widen the detection range. The SAR feature is typically used for electromagnetic wave-based detection because it is the satellite that best satisfies these conditions. This inevitably causes a delay in the detection of oil spills, making it difficult to perform small-scale detection. In contrast, the image-based detection can detect a small range immediately and prevent large accidents in advance.

Thus, in this study, an oil pollution detection system is proposed that uses the deep learning-based algorithm You Only Look Once (YOLO) [5], a tool recently used in the image processing field.

The remainder of this paper is organized as follows. Section 2 briefly describes the YOLO algorithm, which is the technology used in this study. Section 3 describes in detail the structure of the proposed system and Section 4 provides an evaluation of the

<sup>†</sup> Corresponding Author (ORCID: <http://orcid.org/0000-0002-8133-535X>): Research Professor, Department of Control and Automation Engineering, Korea Maritime & Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan 49112, Korea, E-mail: omega@kmou.ac.kr, Tel: 051-410-5294

1 Ph. D. Candidate, Department of Computer Engineering and Interdisciplinary Major of Maritime AI Convergence, Korea Maritime & Ocean University, E-mail: homin2006@hanmail.net, Tel:051-410-4896

2 Ms. D. Candidate, Department of Control & Automation Engineering, Korea Maritime & Ocean University, E-mail: monotic1301@naver.com, Tel: 051-410-4896

3 President, Intelligent System Technology, E-mail: jungkn.ist@gmail.com, Tel: 070-7347-6678

4 Professor, Department of Control and Automation Engineering and Interdisciplinary Major of Maritime AI Convergence, Korea Maritime and Ocean University, E-mail: jhoon@kmou.ac.kr, Tel: 051-410-4574

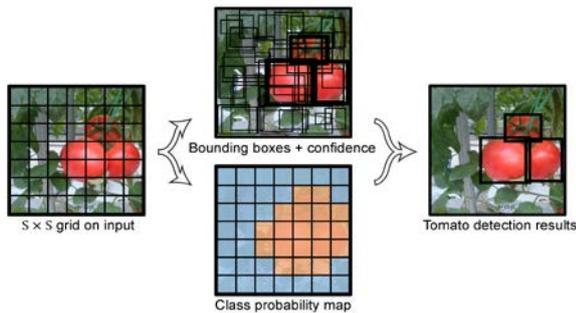
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

performance of the proposed system. Finally, Section 5 concludes the study and discusses future research.

## 2. Related Work

### 2.1 YOLO algorithm

The YOLO algorithm considers the bounding box location and the class possibility on a video or an image as a regression problem and predicts the object type and location by analyzing the image only once. The YOLO algorithm divides the input video or image into  $S \times S$  grids and measures the location of the bounding box that encases the labeled object. The anchor boxes separated by the grid and overlapped on the bounding box estimate the class probability  $p(\text{class} | \text{object})$ , object existence probability  $p(\text{object})$ , object center position  $(x, y)$ , and the width and height of the bounding box (width, height) to measure the confidence score. Consequently, the YOLO network outputs two items: data regarding the bounding box and confidence value. **Figure 1** illustrates the entire process.



**Figure 1:** An example of the YOLO algorithm object detection process

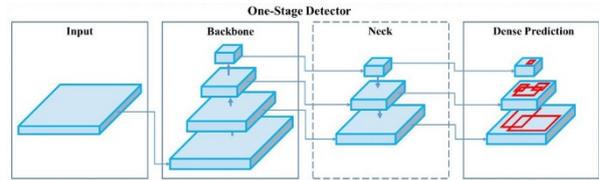
In **Figure 1**, the input image is segmented with a  $7 \times 7$  ( $S = 7$ ) grid to estimate the bounding box and class probability ( $\text{Pr}(\text{class})$ ) and to combine these aspects to recognize the object location and class. The confidence score, as shown in **Equation (1)**, indicates the existence of an object within a bounding box and the extent of the reflection of the corresponding class.

$$\text{confidence score} = \text{Pr}(\text{class}) \times \text{IoU} \quad (1)$$

The intersection over union (IoU) is the value of the size of the intersection between the answer box and prediction box divided by the size of their union. This indicates the extent of the overlap. Thus, when the intersection is nonexistent, the IoU value is equal to zero, whereas a complete overlap results in an IoU value of one.

### 2.2 YOLO v5

YOLO v5 [6] is the latest version of the YOLO algorithm released in 2020. **Figure 2** shows the object detection process using YOLO v5.



**Figure 2:** The object detection process of YOLO v5

The major difference between YOLO v5 and its previous version YOLO v4 [7] is the number of deep learning model parameters, which is one-tenth of that of YOLO v4. This is because YOLO was developed on Tensorflow [8] up to v4 but was implemented on Pytorch [9] in v5. Another difference is the training method used in the deep-learning model. First, data augmentation is automated through scaling, color-space adjustments, and mosaic augmentation. Second, unlike previous versions in which the bounding box sizes are determined using k-means and genetic learning algorithms, YOLO v5 acquires them with deviations in the anchor box dimensions.

### 2.3 Marine oil spill detection

Marine oil spill detection technologies enable the detection of the location, type, and quantity of the oil spill [10]. These are largely classified into two methods of detection: electromagnetic wave-based detection using radar and image-based detection using cameras. The electromagnetic wave-based method determines the corresponding object according to the time it takes for the generated wave to return after its reflection on the object and the changes in wavelength. The image-based method determines an object according to its color, shape, and continuous movement displayed on the recorded images.

Electromagnetic wave-based detection uses data recorded from the synthetic aperture radar (SAR) of a satellite to detect oil spills [11]. Satellite radar transmits microwaves with long wavelengths; thus, data can be obtained regardless of weather and time. As sea surfaces covered with oil reflect most of the microwaves, differences in values are generated in the measured data in comparison to the normal sea surface.

Image-based detection uses deep learning or machine learning models based on recorded image data to detect oil spills [12]. Infrared cameras are mainly used for this process. Objects in

infrared images differ in color depending on their thermal energies, and training data are fabricated by segmenting the images into frame units and labeling the range of oil spills. If a general camera that outputs RGB images is used instead of an infrared camera, there are two drawbacks. First, the difference in temperature between oil spill from a ship and seawater cannot be used as an additional feature. The second issue is the light reflections from oil and seawater, resulting in additional noises in the image. Thus, a real-time system for detecting oil can be structured by employing infrared camera.

The electromagnetic wave-based detection method has two disadvantages as it uses satellites. First, real-time detection is difficult. Second, transmitting microwaves from space to the sea is only practical for large-scale spills such as oil tanker accidents. In contrast, image-based detection requires only an infrared camera and a small amount of embedded equipment for image analysis. An advantage of this method is the feasibility of its installation on a small ship. Therefore, an infrared camera was used in this study to record an artificially fabricated oil spill and build training data for the YOLO v5. The constructed training data were used to train YOLO v5 for use in an oil spill detection system.

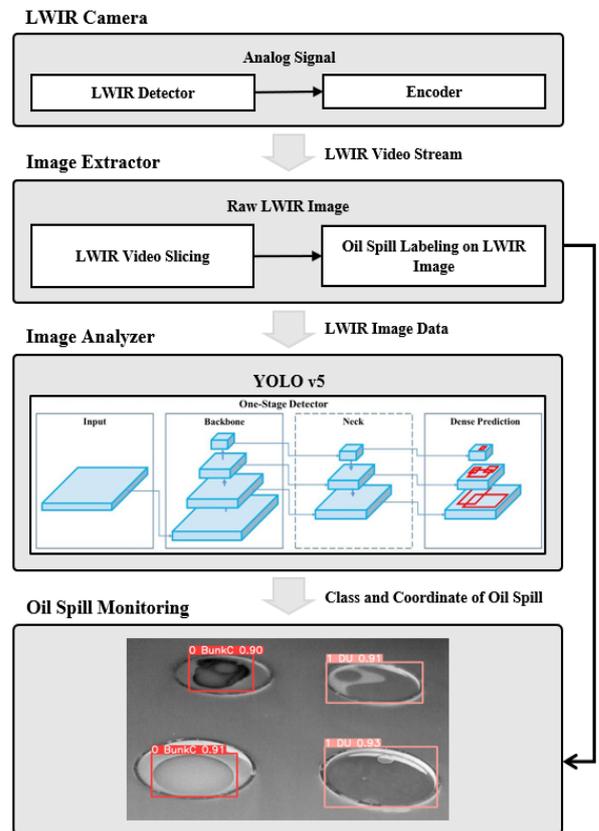
### 3. Oil Spill Detection System with LWIR Camera

In this study, a deep learning-based oil spill detection system using a long-wave infrared (LWIR) camera is proposed, and its flow chart is presented in **Figure 3**. The detection process is divided into four major steps. First, an LWIR camera records thermal images. Second, an image extractor slices the images using frames. Third, an image analyzer detects oil spills from the sliced images. Finally, the outputs of the image analyzer as images and detection results are shown to the users for monitoring. The image analyzer is an oil spill detection system introduced in Section 2, which is based on deep learning using the YOLO v5 model. The image extractor receives infrared images from the LWIR camera and dissects them into frame units that can be used as inputs for the image analyzer. The dissected images additionally go through a labeling process according to the oil location and size such that they could be used as training data. The following sections provide detailed descriptions of the image extractor and image analyzer.

#### 3.1 Infrared imaging

An LWIR camera was used for imaging to train and evaluate the proposed system. Bunker C oil and diesel oil, which are

mainly used as vessel fuels, were selected as oil types for detection. In the recording setting, an acrylic cylinder was inserted into a water tank filled with seawater, and oil was poured on top. An acrylic cylinder was used to eliminate the influence of wind. **Table 1** summarizes the types of images recorded.



**Figure 3:** Flowchart of the deep learning-based oil spill detection system with LWIR camera

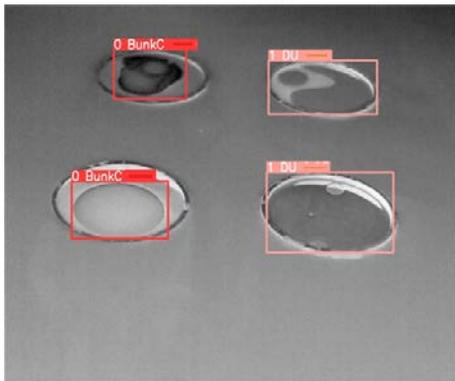
**Table 1:** Types of videos recorded with LWIR camera

Type	Details
Day 1	Recording after pouring oil into cylinder at 10 a.m.
Day 2	Floating matter (leaves and paper pieces) added to the Day-1 condition
Day 3	Recording after pouring oil into cylinder at 1 p.m.
Night 1	Recording after pouring oil into cylinder at 10 p.m.
Night 2	Recording without pouring oil at 10 p.m.
Night 3	Floating matter (leaves and paper pieces) added to the Night 1 condition

#### 3.2 Image extractor

The image extractor receives thermal images from the LWIR camera, segments them into frame units, and converts them into

images for each frame. The proposed extractor segments the images into units of 10 frames. Segmented images used as data for the experiments are immediately delivered to the image analyzer, and those used as training data underwent an additional labeling process. The oil locations are marked on the images and the oil types are recorded. An example of labeling performed on a segmented image is shown in **Figure 4**. Bunker C oil (Bunk C) and diesel oil (DU) labels were attached to each oil mass and poured into the cylinders.



**Figure 4:** An example of labeled training data (video type: Day 1)

The labeled training data were transformed into vertex coordinates in the form shown in **Figure 5**.

0 0.857232 0.437374 0.093284 0.144432

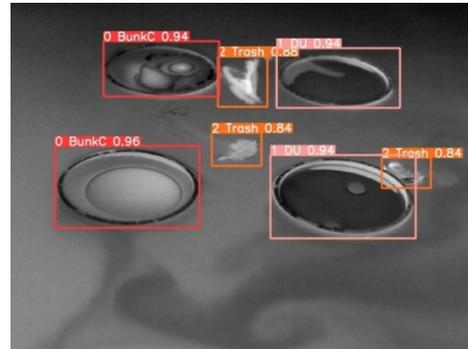
**Figure 5:** An example of transformed training data

Each value represents, from left to right, the class number, bounding box x-coordinate, bounding box y-coordinate, width, and height of the object, and each vertex coordinate is considered as one bounding box and saved as a text file. The bounding box data from each image were used as the training data for the deep learning-based model (YOLO v5). When training was completed by the algorithm described in Section 2, the oil type and location were detected from the input images to obtain the results as outputs.

### 3.3 Image analyzer

The image analyzer receives the dissected thermal images from the image extractor, detects oil spills, and outputs the data. The YOLO v5 deep learning model described in Section 2 was used in the detection process. The image analyzer performs two processes: training and testing. During the training process, labeled training data are received from the image extractor. After

the training, the testing process is performed, where the oil spill is detected with only thermal images, and no labeled images are received as inputs. **Figure 6** shows an example of the oil and floating matter detected on a thermal image using an image analyzer.



**Figure 6:** Example of detection results from an oil pollution detection system (video type: Day 2)

## 4. Experiment and Evaluation

### 4.1 Experimental setup

An LWIR camera was used to record each image type, as shown in **Table 1**. Each recording was performed for 10 minutes. The ‘DarkLabel’ program was used to construct training and testing data from the recorded images. Bunker C oil, diesel oil, and floating material labels, and the x-coordinate, y-coordinate, width, and height values were attached to all acquired images through this program to construct the data. Consequently, 1,644 labeled thermal image data were built. The train: verification: test ratio was set to 7:2:1, in which 1,151, 331, and 162 units comprised each category.

### 4.2 Parameters and structure of deep learning model

As hyperparameters used in the training process, the batch size and subdivision values were both set to 64 and the number of epochs was set to 61,000. The computational resources used for the evaluation are listed in **Table 2**.

**Table 2:** Computational resources used in evaluation

Type	Device name
Operating System	Ubuntu 18.04.1 LTS(GNU/Linux 4.15.0-66generic x86_64)
Processor	Intel® Xeon® CPU E5-1660 v3 @ 3.00GHz
Memory	64Gb
Graphics Feature	TITAN RTX 24GB

CUDA Version	10.1
--------------	------

YOLO v5 darknet53 conv.137 was used as the deep learning model, and the corresponding network structure is shown in **Figure 7**.

Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1
	Convolutional	64	3 × 3
	Residual		128 × 128
	Convolutional	128	3 × 3 / 2
2×	Convolutional	64	1 × 1
	Convolutional	128	3 × 3
	Residual		64 × 64
8×	Convolutional	256	3 × 3 / 2
	Convolutional	128	1 × 1
	Convolutional	256	3 × 3
	Residual		32 × 32
8×	Convolutional	512	3 × 3 / 2
	Convolutional	256	1 × 1
	Convolutional	512	3 × 3
	Residual		16 × 16
4×	Convolutional	1024	3 × 3 / 2
	Convolutional	512	1 × 1
	Convolutional	1024	3 × 3
	Residual		8 × 8
Avgpool		Global	
Connected		1000	
Softmax			

**Figure 7:** YOLO v5 darknet53 conv.137 model architecture

### 4.3 Performance analysis

The training and testing of the deep learning model proceeded using a model similar to that shown in **Figure 7**. **Table 3** summarizes the test results, which are expressed as a confusion matrix.

**Table 3:** Confusion matrix of deep learning-based image analyzer

Confusion Matrix		Actual class			Total
		P0	P1	N	
Predicted class	P0	67	0	1	68
	P1	0	68	1	69
	N	2	1	22	25
Total		70	70	22	162

% P: Positive, N: Negative

TP: True Positive, FP: False Positive

TN: True Negative, FN: False Negative

The images with and without oil spills in **Table 1** were expressed as positive (P) and negative (N). In addition, bunker C

oil and diesel oil class numbers were 0 and 1, respectively. Thus, the true positive (TP) value indicates the frequency of accurate oil spill prediction, and the false positive (FP) value indicates the frequency of inaccurate oil spill prediction. A false negative (FN) indicates the frequency of inaccurate indication of no oil spill, and a true negative (TN) indicates the frequency of accurately predicted results of no oil spill. **Table 4** summarizes the performance of the evaluation results using the resources listed in **Table 3**.

**Table 4:** Deep learning-based image analyzer performance

Measure	Equation	Value
Precision, PPV	$\frac{TP}{TP + FP}$	0.9854
Recall, TNR	$\frac{TP}{TP + FN}$	0.9783
Mean Average Precision, mAP	PPV × TNR	0.9640
F1 score	$\frac{2PPV \times TNR}{PPV + TNR}$	0.9818
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	0.9691
False Alarm Rate	$\frac{FP}{FP + TN}$	0.0833
Specificity	$\frac{TN}{FP + TN}$	0.9167

In **Table 4**, precision is a measure that indicates the accuracy of the prediction of the occurrence of oil spills, and the precision rate of the proposed system was 98.54%. This indicates that most of the prediction results were accurate. Recall is a measure that indicates the degree of accuracy of predicting images with oil spills, which was 97.83% for the proposed system. This value indicated that approximately 2% of the oil spill was not adequately detected, which could be further reduced. The F1 score is the harmonic mean of the precision and recall rates, which was 98.18% for the proposed system. Accuracy is a measure that indicates the degree of accurate prediction of all classes or labels, and was 96.91% for the proposed system. Thus, no significant issues are expected to arise when this system is used as a reference to determine oil leaks from actual vessels. The false alarm rate is the rate of falsely predicted oil spill images in cases where there is no pollution. The proposed system had a false alarm rate of 8.33%, which is not a low value. The specificity, denoted as 1-FPR, is the rate that indicates the accurate prediction of images without oil spills, which was 91.67% for the proposed system. In conclusion, the false-alarm rate and specificity of this system can be further improved. The training and testing of the proposed

system were conducted in highly regulated environments, and a small amount of data were constructed compared with other image processing studies. Therefore, the results are inadequate for conducting field tests. This system requires further supplementary studies.

#### 4.4 Discussion

The cases of previous studies described in Section 2.3 were compared with the method proposed in this study regarding oil spill detection. In the present study and that of De Kerf et al. (2020) [12], image-based detection was performed using convolutional neural network (CNN) [13]-based deep learning models YOLO v5 and MobileNet, respectively [14]. In contrast, Singha et al. (2016) [10] performed an electromagnetic wave-based detection and used ‘support vector machine (SVM)’ [15], the machine learning model. Different models were employed, because the form of the data varied according to the detection method. Image-based detection utilizes sliced recorded images in the detection process. In contrast, electromagnetic wave-based detection utilizes the numerical data of microwaves transmitted and collected by the SAR in the detection process. **Table 5** summarizes the aforementioned methodologies.

**Table 5:** Comparison of methodologies used in this study and cases referenced in Section 2.3

Type	Methodology	Model Used	Accuracy
Fan Li et al., 2015	Electromagnetic wave-based	SAR features, CRF	0.9438
Singha et al., 2016	Electromagnetic wave-based	SAR features, SVM	0.90
De Kerf et al., 2020	Image-based	MobileNet	0.89
Our approach	Image-based	YOLO v5	0.9691

Comparing the accuracy according to the time sequence, it can be observed that our approach has the highest accuracy. In addition, two electromagnetic wave-based methods use a machine learning method for classification, although they show relatively high accuracy. This is expected to further improve the performance when the model is changed to a deep learning-based model.

However, the oil spill detection cannot be conducted fast because it requires the use of features from SAR. Such restrictions have changed the direction of research toward image-based, like ‘MobileNet’ and our approach. ‘MobileNet’ showed relatively low accuracy with 89%, but our approach showed the highest

accuracy with 96.91%. This confirms that oil spills can be efficiently detected with high accuracy and speed without using SAR features.

## 5. Conclusion

In this study, an oil spill detection system is proposed and developed. The system uses images from an LWIR camera and a deep-learning-based YOLO algorithm that operates without specific sensors or equipment. A total of 1,644 images were detected, with an accuracy rate of 96.91% and a false alarm rate of 8.33%. This is by no means inferior compared to other methodologies, but it requires additional performance improvement and expansion of the study purpose. For our experiments, we used a water tank filled with seawater. In addition, to reduce the effect of wind, an acrylic cylinder was used during infrared imaging. This may be different from actual oil spills in terms of size and shape. Therefore, future studies will aim to increase the accuracy of oil spill detection and lower the false alarm rate by not only increasing the amount of data but also by increasing the types of oil spills.

## Acknowledgement

Following are results of a study on the “Leaders in Industry-university Cooperation+” Project, supported by the Ministry of Education and National Research Foundation of Korea.

## Author Contributions

Conceptualization, H. M. Park and S. D. Lee; Methodology, H. M. Park and J. H. Kim; Software, G. S. Park; Validation, G. S. Park and J. H. Kim; Resources, G. S. Park and J. K. Kim; Writing—Original Draft Preparation, H. M. Park; Writing—Review & Editing, J. H. Kim, S. D. Lee and J. K. Kim.

## References

- [1] W. -G. Jhang, J. -H. Nam, and G. -W. Han, “Advancement of land-based pollutant management system,” Korea Maritime Institute, 2012 (in Korean).
- [2] H.-M. Park, J.-H. Kim, “Predicting sentence of ship accident using an attention mechanism based multi-task learning model,” Proceedings of The Korean Institute of Information Scientist and Engineers, pp. 448-450, 2021 (in Korean).
- [3] D. -S. Kim, “Oil spill detection from dual-polarized SAR images using artificial neural network,” Master Thesis,

Department of Geoinformatics, Graduate School of University of Seoul, Korea, 2017 (in Korean).

- [4] H. -M. Park, G. -S. Park, Y. -R. Kim, J. -K. Kim, J. -H. Kim, and S. -D. Lee, "Deep learning-based drone detection with SWIR cameras," *Journal of Advanced Marine Engineering and Technology*, vol. 44, no. 6, pp. 500-506, 2020.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.
- [6] Github, <https://www.github.com/ultralytics/yolov5>, Accessed December 1, 2020.
- [7] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: optimal speed and accuracy of object detection," *arXiv:2004.10934*, 2020.
- [8] Tensorflow, <https://www.tensorflow.org>, Accessed December 1, 2015.
- [9] Pytorch, <https://www.pytorch.org>, Accessed December 1, 2016.
- [10] M. Fingas and C. E. Brown, "A review of oil spill remote sensing," *Sensors*, vol. 18, no. 1, pp. 91-108, 2018.
- [11] S. Singha, R. Ressel, D. Velotto, and S. Lehner, "A combination of traditional and polarimetric features for oil spill detection using TerraSAR-X," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 11, pp. 4979-4990, 2016.
- [12] T. De Kerf, J. Gladines, S. Sels, and S. Vanlanduit, "Oil spill detection using machine learning and infrared images," *Remote Sensing*, vol. 12, no. 24, p. 4090, 2020.
- [13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, 1989.
- [14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv:1704.04861*, 2017.
- [15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.