Leveraging SARs and Advanced Deep Learning Techniques for Oil Spill Detection in UAE

Munya Alkhalifa[a], Aisha Alowais[a], Antonios Manousakis[a,b]

[a] Sharjah Academy for Astronomy, Space Sciences, and Technology, University of Sharjah, Sharjah, 27272, United Arab Emirates
[b] Department of Applied Physics and Astronomy, University of Sharjah, Sharjah, 27272, United Arab Emirates

Keywords: Oil Spills, AI, Deep learning, Image Classification, Computer Vision, Earth Observation.

Abstract

Accidental oil spills are known to reflect negative outcomes on the environment and human health as well as marine life and coastal regions’ economy. In aim to solve this issue, we suggest a system that is designed to detect oil spills on the ocean surface and provide information for taking appropriate measures to contain the spill. Our research focuses on two key maritime regions near the United Arab Emirates: The Arabian Gulf and the Gulf of Oman. To create the dataset, we utilized Sentinel-1 Synthetic Aperture Radar (SAR) images that were pre-processed using SNAP for training and SNAPPY in Python for testing. The system uses an automated Vision Transformer (ViT) as its base for the classification and segmentation of oil spills, which was trained on SAR patches falling under two classes. The step of automating the system involves receiving new data inputs and outputting image segments containing the oil slick without delay. Our proposed approach illustrated a high level of performance compared to other Convolutional Neural Network architectures used in similar cases. The ViT accomplished 0.91 accuracy on unseen data with error of 0.3. We put the model into test on new SAR images. The suggested system will help minimizing the effects of oil spills on the ecosystem, human health, and economic losses in the UAE. We believe this study will mark a breakthrough in the management of oil spills as it seeks to safeguard crucial marine and coastal resources through engaging Artificial Intelligence (AI) with cutting-edge algorithms.

1. Introduction

Marine oil spills are considered as a hazardous problem as they pose danger to species, sea ecosystems, tourism, fishing, and other socio-economic value and welfare. The Arabian Gulf and the Gulf of Oman are considered to be vulnerable areas because they lie within a susceptible region. There are large oil processing activities in these areas involving extraction, transportation and even refining; all of which expose the area to high risk of oil spills. The impact of such events is staggering and includes alterations in the marine and coastal structures and environments, changes in coastal income, and risks to the lives of people, depending on these waters. Thus, oil transport and drilling pose more risks in the UAE, making it compulsory to detect and monitor the occurrence of spills.

The identification of oil spills in marine environments remains a paramount area of concern for environmental conservation and the safety of marine fleets, notably in the Arabian Gulf and Gulf of Oman. Despite advances in remote sensing and environmental monitoring technologies, oil spill detection in the UAE has not fully exploited the potential of Synthetic Aperture Radar (SAR) imagery. SAR sensors, particularly those aboard Sentinel-1 satellites, have the ability to penetrate cloud cover and operate in both day and night conditions, offering unique advantages for monitoring vast sea surfaces – as in the UAE and Oman. In light of current artificial intelligence (AI) and deep learning (DL) advancements, there exist specific promising techniques that may help enhance the effectiveness and accuracy of detecting oil spills from SAR images.

SAR imagery captures the backscattering of the facets of the sea surface, causing oil spillage to appear as dark patches with lower backscatter than water (Muji Susantoro & Sunardjanto, n.d.). Nevertheless, distinguishing such spills is not easy because look-alikes exist—phenomena that create formations similar to the dark patches mentioned earlier, such as organic films or low-wind areas (Nunziata et al., 2018). Prior methods of oil spill detection have employed different algorithms depending on the multiple methods like the CNN and other forms of deep learning methods with the purpose of distinguishing the appearance of oil spills (Krestenitis et al., 2019). For instance, some studies have demonstrated that deep learning is useful for developing high accuracy and efficiency in identifying and categorizing oil slicks in SAR images (Zeng & Wang, 2020).

Thus, we suggest a study using Vision Transformers in this regard, which can be regarded as innovative and prospective due to the following qualities: they work with long-range connections and describe the context of the image. This algorithm demonstrated a high level of success in multiple classification benchmarks. Unlike CNNs, where the components mainly extract local features, ViTs could capture the images in their entirety, which could be advantageous when detecting irregularly shaped oil spills (Cheng et al., 2022). SAR imagery recognition can largely benefit from this capability given that the forms of oil spills can be complex and deformed in multiple ways, thus challenging traditional detection (Prastyani & Basith, 2018). Additionally, ViTs’ application can help improve the existing segmentation methods, enabling boundary definition within oil spills with great clarity and subsequently improving the monitoring process (Chen et al., 2023; Song et al., 2020).

Explaining the necessity of Vision Transformers for oil spill identification within UAE’s seaside regions, we believe that applying this proposed approach will empower environmental surveying and provide an immediate response to address the occurrence of such hazardous events. Since this region is a strategic area in terms of shipments of goods and products and is also prone to oil spillage, this pushes the need to have more advanced detection systems for deployment in the actual field.

In the process of SAR imagery and the development of machine learning algorithms, this study expects to contribute towards the constant striving to safeguard marine environments and ensure safety on such imperative waters within the UAE. Our research highlights the alignment with Sustainable Development Goals (SDGs), particularly SDG 14 (Life Below Water), by contributing to marine conservation efforts, and SDG 9 (Industry, Innovation, and Infrastructure) by advancing technological solutions for environmental monitoring. The primary goal is to develop a reliable detection system that can identify oil spills early, enabling timely interventions and minimizing the environmental, public health, and economic impacts of these incidents in the UAE.

2. Methodology

Vision Transformers (ViTs) are one of the breakthrough methods in the computer vision domain, using the transformer structure initially created for working with texts. In contrast to Conventional Convolutional Neural Networks (CCNNs) that work through convolutional layers, ViTs function with image partitions and image parceling where images are split into non-overlapping patches and perceived as sequences akin to words. Thus, ViTs can have the capability to understand the global context or how other information is related in an image thus making them suitable to be implemented in any vision tasks such as image classification, object detection, and segmentation (Dosovitskiy et al., 2020; Jamil et al., 2023).

The pipeline of ViT consists of four main stages: image splitting, patch and position embeddings, transformer encoder, and a multilayer perceptron (MLP) classification head. The first step involves dividing an image into patches of a certain size. These patches will be converted into vectors that are appropriately arranged and flattened and then linearly transformed into a high-dimensional space. In order for these embeddings to retain spatial information, positional encodings are assigned to them. The Self-attention in ViT works using a multi-head attention mechanism to enable the model to determine the relevance of the relative patches and includes all the necessary features in the out-of-the-image representation (Bakhtiarnia et al., 2021; Qian et al., 2021). The output of the attention layers initially goes through the stacked dense layers present in the feed-forward neural network and the final representations which can be used for classification purposes or other use (He et al., 2021; Khan et al., 2021). An overview of the model is depicted in figure 1.

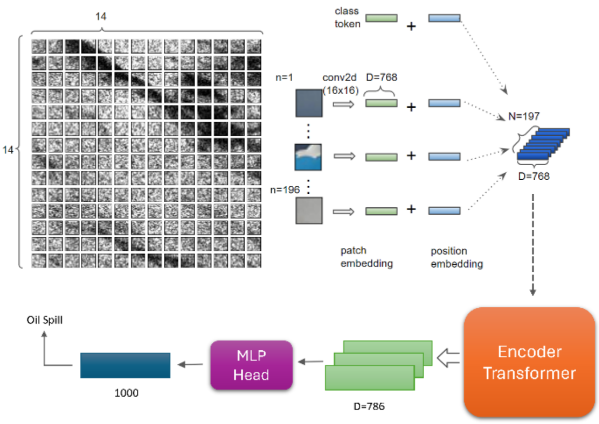

Figure 1. Architecture of the Vision Transformer.

The trained ViT will be the basis for the system that will be receiving inputs from a satellite source and apply segmentation on them. For every new input the classification and localization of oil leaks in the image will be delivered to the operator.

3. Dataset Building

3.1 Data Acquisition

The initial phase involved downloading satellite imagery data from the Copernicus hub and NASA's Earth Data repository from The Earth Science Data Systems (ESDS) program. These SAR images, often acquired from spaceborne sensors on satellites like European Space Agency (ESA) Sentinel-1 satellites, contain valuable information about the surface of the earth. Subsequently, the acquired data was loaded into the Sentinel Application Platform (SNAP), a powerful software tool designed for earth observation tasks. Through this software, we worked on preparing the SAR data by preprocessing and analysis. We exported a total of 16 SAR products from these websites. A SAR product in the software captures very deep and high-resolution features with extensive information, resulting in large file sizes ranging from 900 MBs to 20 GBs. A product is usually visualized in multiple bands, e.g. VV and VH. In our work, we focused on utilizing the amplitude VV polarization mode.

3.2 Data Processing

Pre-processing in image classification is an initial step for preparing the data before deploying them to the model. This will allow further analysis to be enabled after the accuracy is guaranteed and any geometric or radiometric distortions are removed. In this study, we conducted a series of operations to prepare the obtained dataset.

For training, Sentinel-1 SAR data was processed using SNAP version 10.0. extension of SNAP, SNAPPY was used to pre-process the input image before feeding it to the model in Python. SNAP offers free and open-source software for working on data coming from earth observation methods. Through SNAP Desktop, we generate subsets of images from one large SAR image. We conducted a series of operations for every image to obtain a subarea that may or may not contain an oil spill. These phases are:

1. Subset Extraction
   This is the first initial step, which involves extracting a smaller region of interest from the whole SAR image and scanning this area for an oil slick. The more minimized the area, the better the analysis of the region will be, as it facilitates and optimizes the analysis. This helps the oil spill model, integrated with SNAP, focus on more deep features that would not have been clear in large areas.

2. Calibration
   SAR images are generally calibrated into sigma nought, which means performing radiometric calibration to convert digital numbers into a backscatter coefficient.

3. Speckle Filtering
   Radar images are generated through the coherent interaction of the transmitted microwave with the targets, in contrast to optical images. Thus, it is affected by speckle noise due to the cohesive sum interference of the signals scattered from ground scatterers distributed randomly within each pixel (Lee et al., 1994). This noise reduces the image quality when aimed for feature extraction in image classification or segmentation applications. The manifested noise in a radar image is higher than that in an optical image. Sometimes, the speckle noise is minimized by applying

a speckle filter to the digital image before surface display and subsequent analysis.

4. Backscattered Radar Intensity
A radar image is usually represented as a greyscale image. The intensity of each pixel is proportional to the percent of microwave backscattered from an area on the ground depending upon many factors such as type, size, shape, the orientation of scatterers at the target area, moisture content present at the given area, microwave frequency and polarization of radar pulses, and the incident angle of a radar beam. The backscattering coefficient, or normalized radar cross section (dB), is a physical quantity into which the pixel intensity values are converted. The values of this quantity fluctuate from 5 dB for very bright objects up to -40 dB for very dark surfaces. When recognizing oil on the sea surface, usually in a SAR image, a serene ocean surface exhibits dark color while rough surfaces may appear bright, particularly when observed at a small angle. Under certain conditions, when the sea surface seems to be rough, oil films are detected by appearing as dark patches.

5. Oil Spill Detection
Determining the appearance of oil slick in SAR images can be done through three approaches: manual, semi-automatic, and automatic (Topouzelis, 2008). In this study, we follow the automatic approach, which involves techniques such as dark spot detection and RoI identification. Therefore, we make use of SNAP’s oil spill detection application to determine whether the chosen region shows an oil slick.

3.3 Data Augmentation

After analysing 16 SAR products we exported 168 sub-views. 120 of these views contain oil spill, while the other 120 are either do not show a spill or may exhibit something similar to oil. The data obtained was split into two classes, namely, spill and no-spill. Each class consisted of 84 images for training while reserving 36 images for validation, leading to a total of 120 images per class. Furthermore, three augmentations were selected to populate the data, these included cropping, brightness, and rotation. A sample of the performed augmentations is presented in figure 2. Post-augmentation photos resulted in 252 training images, and 108 validation images for each class. This resulted in a total of 360 images per class. Table 1 summarizes dataset size details.

	Number of Images	
	Original	Augmented
Spill Class	120	360
Non-Spill Class	120	360
Total	240	720

Table 1. Dataset size before and after augmentations

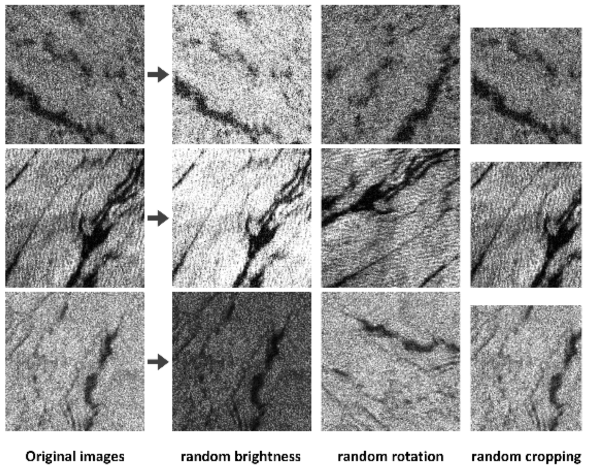

Figure 2. The first column contains original images from the dataset showing oil spills. Each row showcases the applied transformation on the image.

4. Results and discussion

4.1 Model Evaluation

Using python and PyTorch (Paszke et al., 2019) in Google Colab, we trained the Vision Transformer model for 100 epochs. Table 2 shows the resulting evaluation metrics. A validation accuracy of 91% means that the model successfully classifies 91% of the validation dataset, demonstrating how effective the model is in making decisions. However, for better evaluation, other metrics must be considered. For example, is it noticed that precision and recall have same value (0.9), which indicates how well the model is recognizing the true positives and decreasing the false negatives and false positives. F1-score of 0.9 reflects how balanced the model’s performance across precision and recall is, thus showing a harmonic mean between them. Validation loss was calculated with the cross-entropy function resulting in loss of 0.3 which implies that the model is performing better in classifying the true positive. From all previous values, the Visual Transformer proved to achieve robust performance by exhibiting strong values across all metrics hence suggesting how it can generalize unseen new data. To support these finding we also illustrate the confusion matrix in figure 3. Despite that seven no-spill samples were misclassified, the model was able to correctly identify the majority of them, which is 29. Moreover, it succeeded in recognizing 35 true positives out of 36. Overall, the model had minimal errors in recognizing new data.

	Results	
	Training	Validation
Accuracy	0.99	0.91
Loss	0.01	0.30
Precision	0.99	0.90
Recall	0.99	0.90
F1-Score	0.99	0.90

Table 2. Resulted evaluation metrics from training and validation

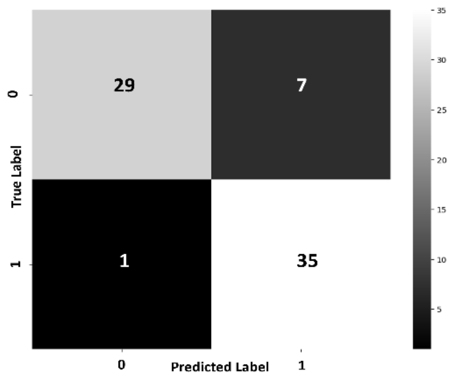

Figure 3. Confusion matrix on validation data.

For a deeper analysis of the model's performance, we observe both accuracy and loss curves during training across 100 epochs, as shown in figure 4. The training loss (blue solid line) is reduced significantly in the initial epochs and remains below 0.05 starting from the epoch number of 25. Validation loss (orange dashed line) fluctuates significantly, especially between epochs 20 to 60, where a spike is observed. This implies that generalization of the model is unstable at these epochs. However, it starts to stabilize after epoch 70 and remains at a higher level than the training loss. Training accuracy (green solid line) climbs steeply and then levels off at around 100% after a few epochs, which shows that the present model is far doing well with the training data set. Validation accuracy (red dashed line) increases gradually at the initial steps and remains oscillatory after 20 epochs, with a value of around 90%. This may indicate an overfitting problem. The plot also addresses issues with overfitting and validation stability during the training. Hence these insights are important to be considered for optimizing the model's performance.

Figure 4. Evolution of accuracy and loss during training and validation.

4.2 Model Inference

To comprehensively assess the performance, we performed model inference with 16 SAR images, in which eight of them were from UAE, and the others were from other countries. These images undergo pre-processing steps before being fed into the ViT. Unlike the training process, where we used small segments showing spill to train the model, the test images represent full SAR images, mimicking the output from satellite acquisitions. Before feeding these images into the model for prediction, similar to previous steps in the pre-processing phase, we used the same pipeline to handle and prepare the test samples using SNAPPY in Python instead of SNAP Desktop.

Each patch was then individually fed into the Vision Transformer model for inference. If any patch was classified as containing an oil spill, the entire SAR image was labelled as an oil spill. Additionally, patches with detected spills were highlighted to provide a visual justification for the classification. Of the eight SAR images from the UAE, six were successfully classified as containing oil spills. Similarly, five of the eight images from other regions were correctly identified. This demonstrates the model's ability to generalize beyond the training dataset, accurately detecting oil spills in diverse environments.

Figure 5 shows an example of one batch undergoing attention mechanisms in the Vision Transformer (ViT) model employed to identify oil spills in synthetic aperture radar (SAR) images. The first one on the extreme left is the input SAR image highlighting regions that share the characteristics of an oil spill. The next eight heatmaps depict the attention maps generated by different attention heads of the Vision Transformer model. Heatmap is the visibility distribution of a particular patch grid of the image, and the sum of all heatmaps provides an attentive evaluation of the image. The heatmaps show the distribution of attention concentration, where yellow color indicates increased attention and blue color – reduced attention. The heat maps depicted in this plot emphasizes that Vision Transformer layers focus on a given area of the picture and hence represent both local and global embeddings. The feature of multiple attention heads allows the model to focus on different areas of an image at once, thus improving its ability to recognize oil spills as well as filtering out similar formations or noise. The proposed multi-headed attention mechanism improves the model's resistance to SAR image complexity and the accuracy of oil spill detection.

Figure 6 displays the results after testing the model on a full SAR image. The original SAR image is shown in (a). In (b), the prediction outputs are displayed. Each patch is an input to the ViT. The model identifies the segments having an oil spill and assigns their class accordingly with a confidence score. Instead of classifying the entire SAR image, the image was divided into small regions of interest. This helped the model to focus on specific areas similar to what was trained on and avoid making false positives when identifying finer details. This patch-based methodology enables the model to precisely detect areas of oil presence, even within intricate SAR images where environmental noise or natural analogues may otherwise conceal spills. The high confidence scores indicate the efficacy of the Vision Transformer model in differentiating oil spills in the SAR images.

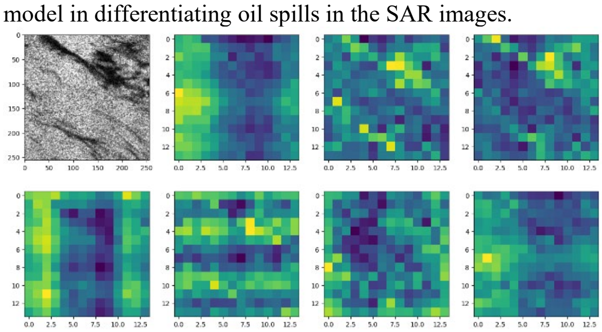

Figure 5. Visualization of attention.

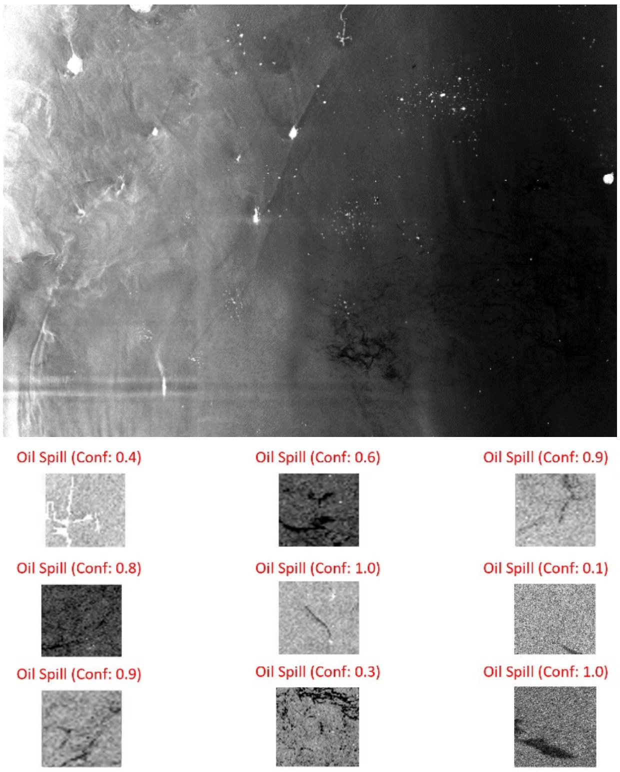

Figure 6. (a) views the input SAR image. in (b) we plot random segments from the original input image displayed after classification. A confidence value will determine the strength level of oil spill assumption. The bigger the confidence the closer the prediction goes towards oil spill class.

5. Conclusion

An automatic oil spill detection system based on ViT's segmentation and classification architecture was suggested in this study. This serves the purpose of taking immediate action against the incident of oil leakages in the ocean. The project focuses on areas surrounding UAE including Arabian Gulf and Gulf of Oman. We trained the model on segments of SAR images obtained from online resources provided by Copernicus Hub and ESA. The resulting model was tested on new SAR data that were segmented and fed to the model as patches. This allowed localizing the leakage that exists in the image. The model showed signs of overfitting which can be solved by enhancing both the dataset and the architecture of the neural network. We believe that our proposed observation system can assist in reducing the risks driven by oil slick appearance which affect the environment, human health, and the country's economy. Following the interest of achieving a sustainable future in the UAE, we aim to improve our current findings using more elevated state-of-art technology.

References

Bakhtiarnia, A., Zhang, Q., & Iosifidis, A. (2021). *Multi-Exit Vision Transformer for Dynamic Inference*. http://arxiv.org/abs/2106.15183

Chen, F., Balzter, H., Zhou, F., Ren, P., & Zhou, H. (2023). DGNet: Distribution Guided Efficient Learning for Oil Spill Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, *61*. https://doi.org/10.1109/TGRS.2023.3240579

Cheng, L., Li, Y., Zhang, X., & Xie, M. (2022). An Analysis of the Optimal Features for Sentinel-1 Oil Spill Datasets Based on an Improved J–M/K-Means Algorithm. *Remote Sensing*, *14*(17). https://doi.org/10.3390/rs14174290

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929

He, J., Chen, J.-N., Liu, S., Kortylewski, A., Yang, C., Bai, Y., & Wang, C. (2021). *TransFG: A Transformer Architecture for Fine-grained Recognition*. http://arxiv.org/abs/2103.07976

Jamil, S., Jalil Piran, M., & Kwon, O. J. (2023). A Comprehensive Survey of Transformers for Computer Vision. In *Drones* (Vol. 7, Issue 5). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/drones7050287

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). *Transformers in Vision: A Survey*. https://doi.org/10.1145/3505244

Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., & Kompatsiaris, I. (2019). Early Identification of Oil Spills in Satellite Images Using Deep CNNs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11295 LNCS*, 424–435. https://doi.org/10.1007/978-3-030-05710-7_35

Lee, J. S., Jurkevich, I., Dewaele, P., Wambacq, P., & Oosterlinck, A. (1994). Speckle filtering of synthetic aperture radar images: a review. *Remote Sensing Reviews*, *8*(4), 313–340. https://doi.org/10.1080/02757259409532206

Muji Susantoro, T., & Sunardjanto, D. (n.d.). *OIL SPILL POLLUTION DETECTION LEMIGAS SCIENTIFIC CONTRIBUTIONS OIL SPILL POLLUTION DETECTION USING PALSAR DATA IN TIMOR SEA*. www.guardian.co.uk

Nunziata, F., Buono, A., & Migliaccio, M. (2018). COSMO-SkyMed synthetic aperture radar data to observe the Deepwater Horizon oil spill. *Sustainability (Switzerland)*, *10*(10). https://doi.org/10.3390/su10103599

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. http://arxiv.org/abs/1912.01703

Prastyani, R., & Basith, A. (2018). Utilisation of Sentinel-1 SAR Imagery for Oil Spill Mapping: A Case Study of Balikpapan Bay Oil Spill. *JGISE: Journal of Geospatial Information Science and Engineering*, *1*(1). https://doi.org/10.22146/jgise.38533

Qian, S., Shao, H., Zhu, Y., Li, M., & Jia, J. (2021). *Blending Anti-Aliasing into Vision Transformer*. http://arxiv.org/abs/2110.15156

Song, D., Zhen, Z., Wang, B., Li, X., Gao, L., Wang, N., Xie, T., & Zhang, T. (2020). A Novel Marine Oil Spillage Identification Scheme Based on Convolution Neural Network Feature Extraction from Fully Polarimetric SAR Imagery. *IEEE Access*, *8*, 59801–59820. https://doi.org/10.1109/ACCESS.2020.2979219

Topouzelis, K. N. (2008). Oil spill detection by SAR images: Dark formation detection, feature extraction and classification algorithms. In *Sensors* (Vol. 8, Issue 10, pp. 6642–6659). https://doi.org/10.3390/s8106642

Zeng, K., & Wang, Y. (2020). A deep convolutional neural network for oil spill detection from spaceborne SAR images. *Remote Sensing*, *12*(6). https://doi.org/10.3390/rs12061015