



Oilspillnets: detecting and estimating oil spills using fusion

Besma Guesmi¹ · David Moloney¹

Received: 28 January 2025 / Accepted: 29 July 2025 / Published online: 18 August 2025
© Springer Nature Switzerland AG 2025

Abstract

Oil pollution poses a significant environmental threat to oceanic and coastal ecosystems, with severe consequences for marine biodiversity, water quality, and local economies. The recent sinking of the MT Terranova tanker in Manila Bay (25 July 2024) released 1.5 million liters of industrial oil into the sea, exacerbating ongoing spill risks. Previous large-scale incidents (e.g., Deepwater Horizon, 2010; Exxon Valdez, 1989) have demonstrated catastrophic ecological and economic impacts, underscoring the need for rapid detection and monitoring. In this work, we propose an end-to-end oil spill monitoring system, "OilSpillNets". Our approach employs a late-fusion (ensemble) strategy that integrates heterogeneous satellite imagery with high-resolution drone data. By capturing large-scale spill patterns via Synthetic Aperture Radar (SAR) and finer details (including oil thickness estimation) from drone imagery, the system provides a comprehensive view of the extent and characteristics of the slick. Lightweight deep learning models are trained on this multimodal dataset and optimised for edge deployment. Evaluation on low-power platforms (Intel Movidius Myriad X and Rockchip RK3588S) demonstrates real-time performance, with the RK3588S achieving 76 FPS at a power consumption of 3.2 W. These results indicate that our framework can be deployed cost-effectively on satellites, drones, or surface vessels to provide timely early warnings and support cleanup efforts.

Keywords Oil spill · Marine pollution · Deep learning · Remote sensing · Fusion · Detection · Thickness · Volume estimation

Introduction

Large-scale spills like the Gulf War oil spill (1991), in which an estimated 380 to 520 million gallons of oil spilled (Statista 2023). The BP Deepwater Horizon catastrophe in 2010 spilled 205.8 million gallons of oil and 225,000 tons of methane into the Gulf of Mexico and had a devastating environmental impact on the marine ecosystem and coastal habitats, as extensively studied in Beyer et al. (2016) and Readman et al. (1996). According to the latest oil tanker spill statistics report (International Tanker Owners Pollution Federation 2019), 66% of the registered oil spills are

medium-sized (2,155–215,502 gallons), and 53% of those oil spills occur inside ports (De Kerf et al. 2020). Oil spills appear as dark regions with distinct boundaries on the water surface when observed in SAR imagery due to their capacity to reduce the backscattered energy as it dampens small-scale surface capillaries and short gravity waves. Identifying the boundaries of oil spills enables continuous monitoring of their spread over time and determination of their size. However, the appearance of oil spills in SAR images may differ due to variations in the radiometric characteristics of radar imaging, which depend on factors such as wavelengths, frequencies, and polarisations (Al-Ruzouq et al. 2020).

Conventional approaches to identifying oceanic phenomena in SAR imagery rely heavily on hand-crafted features and predefined detection thresholds, leading to poor generalisation and susceptibility to noise interference. However, recent research has shown that deep learning has revolutionised this field. Using deep learning algorithms such as convolutional neural networks (CNNs), SAR imagery can be analysed with greater precision and adaptability. Deep learning models autonomously learn complex patterns indicative

Responsible Editor: Riadh Ahmadi.

✉ Besma Guesmi
besma.guesmi@ubotica.com

David Moloney
david.moloney@ubotica.com

¹ Ubotica Technologies, DCU Alpha, Old Finglas Road, Dublin D11 KXN4, Ireland

of oil spills, eliminating the need for manual feature engineering and fixed thresholds. This advancement enhances efficiency and accuracy in oceanic phenomenon detection, empowering better decision-making and resource management in marine environments. Nevertheless, most existing deep learning–based methods focus exclusively on single-modality SAR data and do not exploit complementary optical or drone data.

Janati et al. (2020) introduced a study on oil spreading, indicating that oil spills can spread over 2 km within as little as one hour, based on the function of oil spill movement speed, which ranges from 0.4 to 0.75 cm/s. This underlines the importance of receiving early warnings to limit the impact of such spill events. Developing a network for rapid detection, tracking, and monitoring of the spread of oil spills enables alerts to be sent and intervention to begin at an early stage. However, to our knowledge, there remains a lack of frameworks that integrate both SAR and high-resolution drone data for property estimation (e.g., thickness), leaving a critical gap in end-to-end spill quantification.

Related work

The following studies represent key advances in SAR-based oil spill detection. De Kerf et al. (2020) proposed an approach for detecting oil spills inside a port environment using only thermal infrared images, utilising CNN, achieving an accuracy of 89%. Mahmoud et al. (2023) proposed the DAM-UNet method, an enhanced U-Net architecture that integrates a dual attention model (DAM) for SAR image segmentation, reaching an accuracy of 94.2% and outperforming many baselines, including traditional U-Net and PSPNet. Yan et al. (2019) proposed the fusion of multilevel features and multiscale features extracted using ResNet50 and Atrous Spatial Pyramid Pooling (ASPP) (Chen et al. 2018), respectively, to detect ocean phenomena in SAR images, achieving an accuracy of 91%. Bianchi et al. (2020) introduced a deep learning framework to detect and categorise oil spills by first performing pixel-level segmentation to detect the presence and then extracting characteristics (shape, texture, contrast). All of these approaches focus on SAR-only pipelines, leaving complementary modalities underexplored.

Recent studies employing deep learning methods demonstrate enhanced performance in detecting and monitoring oil spills. Nonetheless, deep learning models need large amounts of data for supervised learning. Consequently, many oil spill studies use the Oil Spill Detection Dataset developed by Krestenitis et al. (2019). In this study, the data publicly available from Ahmed et al. (2023) are utilised for the detection of oil spills in SAR images. Recently, RGB drone data has been publicly published by De Kerf et al. (2020), which is dedicated to oil spill segmentation,

categorising images into three classes: water, oil, and other. However, the data is not suitable for our pipeline, as the aim of using RGB drone images is to extract oil properties and estimate them. A second RGB drone dataset (Lynn 2023) is publicly available on an online computer vision framework (Ciaglia et al. 2022) and is used in our pipeline as it aligns with our goals. Fingas (2021) introduced the use of the visual appearance of oil spills as a predictor of the thickness of thin oil slicks.

However, while SAR-based models excel at detecting and tracking oil slick extents in all-weather conditions, they cannot estimate physical properties such as thickness. Conversely, RGB drone imagery can infer thickness but lacks large-scale coverage. A unified fusion framework combining SAR and RGB drone data for both detection and property estimation is still missing.

This paper aims to build an end-to-end framework for oil spill monitoring and property analysis using SAR and RGB drone imagery. By applying a late-fusion decision-level ensemble, our approach integrates SAR-based large-scale spill detection with drone-based thickness estimation, a novelty not addressed in prior work. Separate models are trained on SAR and RGB drone data, followed by inference on each dataset. The results are then fused at the decision level to achieve accurate oil spill detection and provide a comprehensive analysis of properties such as thickness, enabling the effective calculation of the oil spill volume and sending alerts/warnings for efficient monitoring. This fusion of modalities and focus on real-time edge deployment distinguishes our contribution.

We then outline the proposed end-to-end workflow for oil spill monitoring (Sect. “[Related work](#)”), detailing the methods used for oil spill segmentation and detection. Section “[Methodology](#)” presents the experimental results, discussing the experiments carried out and their corresponding findings. Section “[Experimental results](#)” concludes the work, including work limitations and future contributions.

Methodology

OilSpillNets

The proposed end-to-end framework, OilSpillNets, seamlessly integrates segmentation models for SAR and RGB drone imagery. Initially, both RGB drone frames and SAR images are processed through a simple and lightweight CNN architecture for a binary classification task. Each network is trained separately on its specific data to effectively filter out frames that contain oil spills from those that do not. Once filtering is complete, the frames containing oil spills are sent to segmentation networks for further processing and analysis. The SAR segmentation model is designed to accurately

delineate the oil spill region, facilitating optimised monitoring and tracking of the spill spread over time. Meanwhile, RGB drone imagery provides critical information about the oil spread limits and thickness estimation through instance segmentation. The information derived from both the SAR and RGB drone models is fused to enhance the extraction of details regarding the detected oil spill regions and their properties. This fusion process enables the precise location of the oil spill and the accurate computation of its thickness, improving the overall analysis of the boundaries and area of the oil spill. Based on the estimated thickness from the RGB drone imagery and the area derived from the fusion of both inputs, the framework can calculate the total volume of the oil spill. This estimation allows for proactive measures, such as sending alerts and warnings when the volume exceeds specific thresholds. For example, if the calculated volume exceeds 264 US gallons (approximately 1000 ls), automated alerts can be generated to notify relevant authorities for immediate response, thus enhancing the effectiveness of spill management and mitigation efforts. Figure 1 highlights the OilSpillNets workflow.

Dataset

In this work, two datasets are used for oil spill detection and property extraction. The first dataset comprises SAR images along with corresponding oil spill masks, which were produced and annotated by De Kerf et al. (2020) and are used for oil spill segmentation. The second dataset includes RGB drone imagery (Lynn 2023), which is used to extract visual properties such as oil spill boundaries and thickness. Detailed descriptions of the datasets can be found in the respective references.

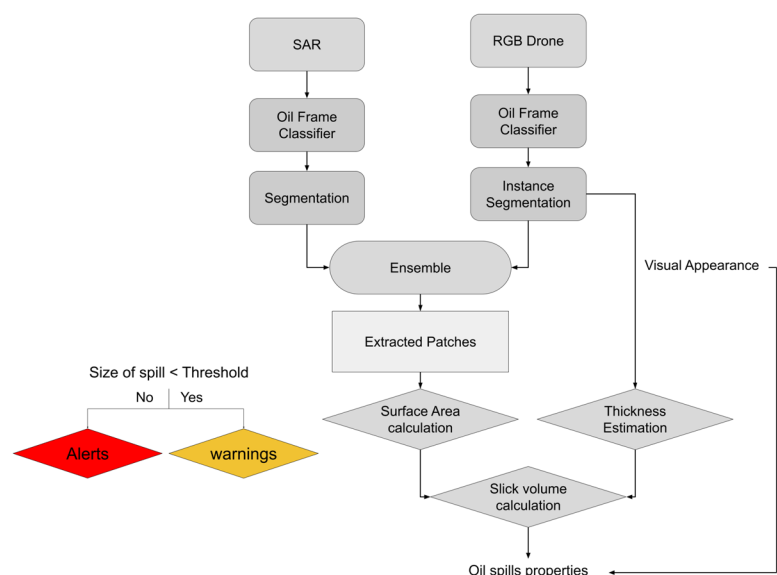
SAR dataset

The chosen data area of study is the northern Mediterranean coast of Egypt, including Alexandria, Rashid, Baltim, Dami-etta, and Port Said cities (De Kerf et al. 2024). The dataset consists of 691 labelled patches of size $256 \times 256 \times 1$, generated from 50 raw SAR images and processed through a series of post-processing steps. The dataset is split into 431 patches for training, 185 for validation, and 75 for testing, and is specifically used for oil spill segmentation tasks. Despite its utility, the SAR dataset size (691 patches) remains modest for deep learning, which may limit model generalisation. Additionally, there is a class imbalance between "oil-spill" and "no-spill" patches, potentially biasing the model toward the majority class. To mitigate this, we apply random under-sampling of "no-spill" patches and augment the remaining data (random rotations $\pm 15^\circ$, scaling 0.9–1.1 $\times$, and Gaussian noise injection $\sigma = 0.01$ –0.05) to increase diversity and reduce overfitting.

RGB drone dataset

The RGB drone dataset (Ciaglia et al. 2022) consists of 148 original images, which are augmented using horizontal and vertical flips, expanding the dataset to 296 images. The augmented dataset is split into 85% for training, 9% for validation, and 5% for testing. This dataset is dedicated to image segmentation tasks aimed at analysing the thickness of detected oil spills. The images are labelled into three thickness categories: Rainbow ($0.3\mu\text{m}$ – $0.5\mu\text{m}$), Sheen ($0.04\mu\text{m}$ – $0.3\mu\text{m}$), and Continuous True Color ($>200\mu\text{m}$) (Bonn Agreement 2016).

Fig. 1 OilspillNets framework



Oil spill detection

For SAR segmentation and model training, we selected several pretrained backbones to balance accuracy and inference speed. Specifically, we evaluate FCN (Fully Convolutional Network) (Long et al. 2015) with ResNet50 or ResNet101 backbones, DeepLabV3 (Chen et al. 2018) with ResNet50 or ResNet101, and DeepLabV3 with MobileNetV3 Large. FCN provides a well-established semantic segmentation baseline, while DeepLabV3's atrous spatial pyramid pooling leverages multi-scale context, which is beneficial for variable spill shapes. ResNet50 and ResNet101 backbones offer high representational capacity for maximising accuracy, whereas MobileNetV3 Large is lightweight and enables faster inference on resource-constrained devices. To explore the trade-off between fine-tuning and training efficiency, we also experimented with freezing the pretrained backbone weights, training only the decoder head, which reduces the number of trainable parameters, speeds up convergence on limited SAR data, and mitigates overfitting. However, fully fine-tuning the entire network yielded higher segmentation accuracy at the expense of longer training time.

In parallel, the RGB drone data is classified and segmented using YOLOv8 (Reis et al. 2023) instance segmentation with the small architecture. YOLOv8 (small) was chosen because its integrated instance segmentation head provides precise thickness categorisation while maintaining low inference latency, making it suitable for real-time deployment on edge platforms. By comparing lightweight backbones (e.g., MobileNetV3, YOLOv8-small) against heavier models (e.g., ResNet101), we identify optimal configurations that satisfy both performance and on-device constraints.

Features extraction

The feature extraction of the oil spill monitoring leverages the strengths of both SAR and RGB drone imagery to provide a comprehensive analysis of the detected oil spill regions and their properties, such as thickness and volume estimation.

SAR imagery for oil spill detection

Due to its ability to capture detailed surface information regardless of lighting or weather conditions, SAR imagery is used to accurately detect and extract oil spill regions by training a segmentation model to extract the boundaries of the oil slicks. The segmentation of SAR data provides key spatial information that is critical for calculating and extracting the region of the detected spills. These extracted regions

are used to monitor the spread of oil accurately over time and are combined with thickness estimates from RGB drone imagery, allowing for precise oil spill volume calculations.

RGB drone imagery for oil spill thickness

The thickness of an oil slick on water is closely related to its color appearance due to various physical processes such as reflection and absorption of light. Thin oil slicks, defined as those below $3\text{ }\mu\text{m}$, have higher reflectivity because they absorb less light, resulting in bright, visible colors like rainbow or sheen patterns. As the oil slick thickens, light absorption increases, leading to darker or more muted colors, such as continuous true color, which is typically seen in slicks over $200\text{ }\mu\text{m}$ thick. These color-based variations in light behavior, caused by refraction, reflection, and interference at the oil-air and oil-water interfaces, enable estimation of oil slick thickness from visual appearance. Thus, RGB drone imagery is used, in addition to SAR data, to provide high-resolution visual information about the oil spill. The drone data allows for more precise boundary detection and thickness estimation through color-based analysis. RGB images are labelled into different thickness categories based on the observed color patterns (Fingas 2021). For instance, thinner slicks like Sheen ($0.04\text{ }\mu\text{m}$ – $0.3\text{ }\mu\text{m}$) and Rainbow ($0.3\text{ }\mu\text{m}$ – $0.5\text{ }\mu\text{m}$) appear with specific color characteristics, while thicker spills (Continuous True Color, $>200\text{ }\mu\text{m}$) absorb more light, resulting in distinct visual appearances. These thickness levels are vital for calculating the overall volume of the oil spill.

Late fusion

We implement a straightforward two-stage fusion in which SAR imagery first detects and delineates the oil-spill region under all-weather conditions, producing a binary mask of spill extent. That mask is then overlaid onto high-resolution drone images, and within each detected region, we extract visual features, primarily colour-based categories (“sheen,” “rainbow,” “true colour”), to infer oil thickness. In other words, SAR localises the slick footprint, and the drone imagery supplies fine-scale property estimates. By combining these outputs at the decision level, associating each SAR, identified pixel with its corresponding drone-derived thickness, we capitalise on SAR's broad-area sensitivity and drone colour cues for detailed characterisation, all without ever blending raw pixel arrays.

Volume calculation and alerts/warnings

To determine the oil slick volume, the surface area of the detected oil spill, as identified through SAR and RGB drone imagery, is used along with the oil spill edges defined by

instance segmentation boundaries. Next, the estimated thickness, based on the visual appearance of the oil, is applied. The volume of the oil slick (in gallons) is then calculated by multiplying the area of the oil spill by the estimated thickness, where 1 cubic meter is equal to 264.17 US gallons. Figure 2 illustrates the relationship between the oil slick volume (in gallons) and the GSD. Generally, a smaller GSD (i.e., higher resolution) allows for more accurate detection of smaller oil slicks, thereby improving overall volume estimation. This estimated volume is crucial for assessing environmental impacts and determining the necessary response efforts. The formula for calculating the oil slick volume in US gallons is as follows:

$$\text{Total area (pixels)} \times (10 \text{ meters})^2 \times \text{Thickness } (\mu\text{m}) \times 264.17 \quad (1)$$

(1) multiplies the detected area of the oil spill (converted from pixels to square meters) by the estimated thickness (in micrometers) and a conversion factor (264.17) to yield the volume in gallons. The system can also trigger alerts and warnings when the spill volume exceeds predefined thresholds. For instance, if the calculated volume surpasses 500 US gallons, a warning is issued, prompting rapid intervention and mitigation measures.

Low-power device platforms

The implementation of low-power device platforms is essential for effective monitoring and assessment of oil spills, particularly in remote and resource-constrained areas. By enabling real-time data processing and analysis, these platforms facilitate rapid response actions that are critical for mitigating the environmental impact of spills. This technological advancement enhances situational awareness, optimises resource allocation, and improves the overall efficiency of spill response efforts, ensuring timely interventions for both immediate and long-term consequences. Platforms such as the IntelMovidius Myriad X (Petrongonas et al. 2021) and OrangePi RK3588S (Rockchip 2021) are particularly well-suited for deployment in satellite and

drone applications, making them ideal choices for testing and evaluating OilSpillNets.

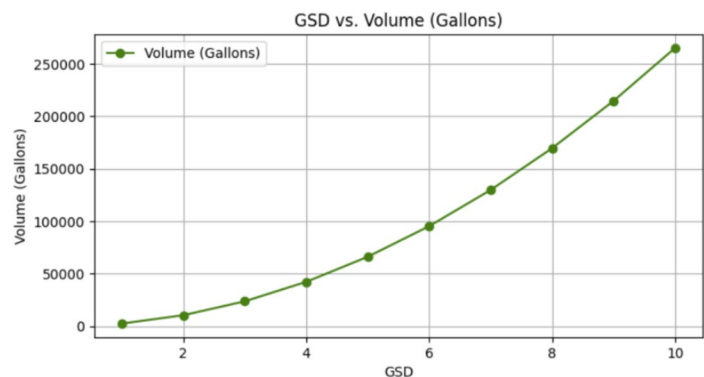
Intel movidius myriadx

The Intel Movidius MyriadX is a cutting-edge AI accelerator specifically featured for accelerating neural network inference. It boasts 16 SHAVE (Streaming Hybrid Architecture Vector Engine) cores and two Neural Compute Engines (NCEs), which operate in parallel to efficiently execute AI and deep learning workloads. One of MyriadX's notable features is its support for two LEON4 processors, each responsible for managing a distinct inference flow via an associated NCE. This configuration allows for the simultaneous processing of two inference tasks, significantly maximising throughput, especially when utilising 8 SHAVE cores per executor. This architecture is particularly advantageous for optimising performance across various neural network models. Additionally, the IntelMovidius MyriadX supports 16-bit floating-point (FP16) data types, which allows for flexibility in precision requirements during neural network computations. This adaptability makes it suitable for a diverse range of applications, from computer vision to robotics, where efficiency and speed are paramount. Its compact form factor and low power consumption also make it ideal for deployment in edge devices, enabling real-time inference in resource-constrained environments.

RockChip RK3588S

The RockChip RK3588S platform presents a robust and versatile configuration, featuring an octa-core CPU that combines high-performance and power-efficient cores. It includes a quad-core Cortex-A76 for demanding tasks and a quad-core Cortex-A55 for energy-efficient processing. This hybrid architecture enables the RK3588S to balance performance with power consumption effectively, making it suitable for a wide range of applications, including edge computing and smart devices. Moreover, the RK3588S features a powerful Neural Processing Unit (NPU) capable

Fig. 2 GSD (Ground Sample Distance) vs. Volume (gallons) of image size 256x256



of delivering up to 6 TOPs (Tera Operations Per Second). This NPU is optimised for accelerating the inference of neural network models and supports a variety of data types, including INT4, INT8, INT16, and FP16. This flexibility enhances the speed and efficiency of model inference, allowing for rapid execution of complex AI tasks. The RK3588S's combination of high-performance computing and dedicated AI acceleration makes it an excellent choice for applications requiring real-time data processing and intelligent decision-making.

Evaluation metrics

Intersection over union (IoU)

Intersection over Union (IoU) is a fundamental metric used in segmentation tasks to quantify the overlap between the predicted segmentation mask and the ground truth mask. It is computed as:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (2)$$

where:

- *TP* (True Positives): Number of pixels correctly classified as positive (Oil spill region).
- *FP* (False Positives): Number of pixels incorrectly classified as positive (not part of the oil spill region).
- *FN* (False Negatives): Number of pixels incorrectly classified as negative (part of the oil spill region but missed by the prediction).

The IoU value ranges from 0 to 1, with higher values indicating better segmentation accuracy. The average IoU across the testing set summarises the model's performance in segmenting oil spill regions.

Dice coefficient

The Dice Coefficient is another important measure of similarity between two sets and is defined as:

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3)$$

This metric also ranges from 0 to 1, emphasising the overlap between predicted and actual sets. A higher Dice score reflects better performance, particularly useful in imbalanced classes, as it gives more weight to the correctly predicted regions.

Mean average precision (mAP)

For RGB drone imagery, Mean Average Precision (mAP) is utilised to evaluate object detection performance, specifically measuring precision and recall across different Intersection over Union thresholds. It is expressed as:

$$\text{mAP}_{50} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (4)$$

where N is the number of classes and AP_i is the Average Precision for class i . Average Precision is calculated by integrating the precision-recall curve for that class. A higher mAP indicates improved detection and localisation of oil spills, critical for effective response efforts.

Inference time

The Inference Time measures the duration (in milliseconds) required for the model to predict training. It provides insights into the efficiency of real-time applications in oil spill response. Low inference times are essential for timely decision-making in emergencies.

Inference per watt (IPW)

Lastly, the Inference Per Watt (IPW) metric evaluates energy efficiency during inference, calculated as:

$$\text{IPW} = \frac{\text{Number of Inferences}}{\text{Power Consumption (Watts)}} \quad (5)$$

Higher IPW values indicate better energy efficiency, which is essential for deploying models on low-power devices, particularly in remote areas for effective oil spill monitoring.

Experimental results

The OilSpillNets framework is evaluated by first assessing the performance of each segmentation network individually, along with the volume estimation of the detected oil spill regions. To comprehensively evaluate the performance of the segmentation networks for both SAR and RGB drone imagery, several metrics are employed.

Oil spill detection

The detection and monitoring of oil spill regions, as well as the control of their spread, are achieved using

segmentation networks applied to two different datasets: RGB drone imagery and SAR imagery.

SAR imagery – network segmentation

Our SAR segmentation pipeline comprises a pretrained feature-extraction backbone (ResNet50, ResNet101, or MobileNetV3-Large), an upsampling decoder, and a binary segmentation head. Models are trained for 100 epochs (learning rate = 0.001, batch size = 16) with early stopping (patience = 10) using binary cross-entropy loss. We evaluate each backbone, both frozen and unfrozen, to assess the trade-off between fine-tuning accuracy and reduced inference cost.

Table 1 reports the key performance indicators (average IoU, average Dice, and per-image inference time) over 67 SAR test images. The best quantitative performance is achieved by DeepLabV3–MobileNetV3 Large (unfrozen), with Avg IoU = 0.9443 and Avg Dice = 0.9711 at 0.0338 s per image. Freezing the backbone yields slightly faster inference but 0.01–0.02 lower IoU and Dice scores.

Figure 3 provides a side-by-side comparison of segmentation outputs on test sample #0 across all backbones and training modes. Panels (a)–(b) compare DeepLabV3–MobileNetV3 Large with frozen versus fine-tuned backbones; (c)–(d) show DeepLabV3–ResNet101 frozen and fine-tuned; (e)–(f) show DeepLabV3–ResNet50 frozen and fine-tuned; and (g)–(j) present FCN–ResNet101 and FCN–ResNet50 in both frozen and fine-tuned variants. In each subfigure, the model’s predicted mask (white) aligns closely with the ground-truth spill region (black), demonstrating that even the lightweight MobileNetV3 and frozen-backbone configurations can capture fine spill boundaries with fidelity consistent with the quantitative IoU and Dice metrics.

Additional qualitative examples, including three challenging cases that clearly illustrate DeepLabV3’s superior boundary delineation compared to FCN backbones, are

provided in Appendix A. These additional examples (see Appendix A, Fig. A.1) include: Sample #1 under low-contrast conditions, where the oil-spill boundary is scarcely distinguishable from ambient sea clutter; Sample #2 highlighting complex shoreline geometry, with a highly irregular coast–spill interface; and Sample #3 demonstrating varying slick thickness, where thinner “sheen” patches sit adjacent to denser accumulations. In each case, DeepLabV3 variants capture finer boundary detail and produce fewer false positives than FCN backbones, in line with the quantitative IoU/Dice improvements reported in Table 1.

RGB drone imagery - YOLOv8 instance segmentation

The RGB drone instance segmentation model is trained using the YOLOv8 small architecture and fine-tuned with the pretrained YOLOv8s model trained on the COCO dataset. The model is trained over 97 epochs with default YOLOv8 hyperparameters and an image size of 256×256 pixels. Table 2 presents the performance of the trained model in classifying and segmenting three different oil spill classes, evaluated across 67 instances.

Despite strong performance on Rainbow and True colour, the Sheen class exhibits substantially lower mAP (0.534 box, 0.437 mask). This imbalance likely arises from (1) fewer Sheen examples during training, (2) low contrast in Sheen regions making them harder to distinguish, and (3) visual similarity between thin Sheen and adjacent water textures. As shown in Table 2, the confusion matrix confirms that Sheen samples are frequently misclassified as Rainbow or True color, and class-wise precision/recall (0.55 each for Sheen versus 0.88/0.96 for Rainbow and 0.90/0.93 for True colour) highlights this difficulty. These results motivate future work to collect more Sheen examples or apply specialised contrast enhancement pre-processing to improve Sheen detection.

The RGB drone imagery instance segmentation model delivered an impressive performance, achieving a mean Average Precision (mAP) of 79.4 % for bounding boxes and 75.8 % for masks (all classes), despite the utilisation of a relatively small labelled dataset.

Oil spill volume estimation

The volume calculation of the detected oil slick regions is based on the extracted information fused from both SAR and RGB drone imagery. The estimated thickness is determined based on visual appearance information from the RGB drone data and the spatial resolution of 10 m for the SAR imagery. Thus, the volume formula (1) is used to compute the selected fused areas presented in Table 3.

Table 1 KPIs (Key Performance Indicators) for proposed SAR segmentation backbones

Network	Freeze	Avg IoU	Avg Dice	Time (s)
FCN-ResNet101	Yes	0.9380	0.9676	0.0079
FCN-ResNet101	No	0.9407	0.9691	0.0139
FCN-ResNet50	Yes	0.7112	0.8244	0.0245
FCN-ResNet50	No	0.6983	0.8150	0.0277
DeepLabV3-MobileNetV3 Large	No	0.9443	0.9711	0.0338
DeepLabV3-MobileNetV3 Large	Yes	0.9337	0.9653	0.0273
DeepLabV3-ResNet50	No	0.9385	0.9679	0.0271
DeepLabV3-ResNet50	Yes	0.9320	0.9644	0.0320
DeepLabV3-ResNet101	No	0.9368	0.9670	0.0273
DeepLabV3-ResNet101	Yes	0.9241	0.9601	0.0551

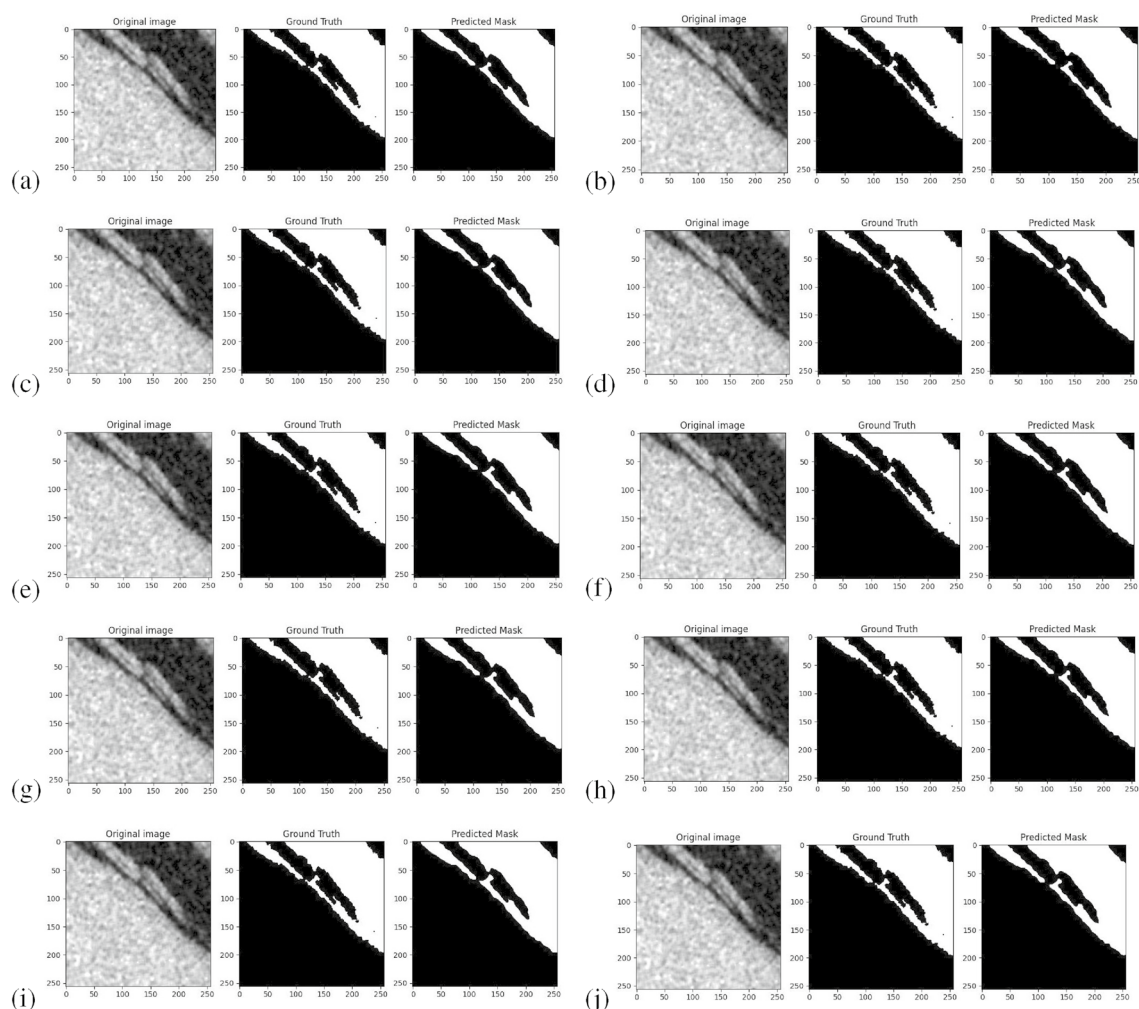


Fig. 3 Qualitative segmentation on test sample #0 for each backbone and training mode: **a** DeepLabV3-MobileNetV3 Large, frozen backbone; **b** DeepLabV3-MobileNetV3 Large, fine-tuned backbone; **c** DeepLabV3-ResNet101, frozen backbone; **d** DeepLabV3-ResNet101,

fine-tuned backbone; **e** DeepLabV3-ResNet50, frozen backbone; **f** DeepLabV3-ResNet50, fine-tuned backbone; **g** FCN-ResNet101, frozen backbone; **h** FCN-ResNet101, fine-tuned backbone; **i** FCN-ResNet50, fine-tuned backbone; **j** FCN-ResNet50, frozen backbone

Table 2 RGB drone instance segmentation metrics: mAP, precision/recall, and confusion matrix

Class	Box (mAP50)	Mask (mAP50)	Precision	Recall
Rainbow	0.995	0.995	0.88	0.96
Sheen	0.534	0.437	0.55	0.55
True color	0.853	0.841	0.90	0.93
Confusion matrix (Actual \ Predicted)				
	Rainbow	Sheen	True color	
Rainbow	22	1	0	
Sheen	3	12	3	
True color	0	2	27	

Table 3 Oil slick volume calculation

Oil spill class	Area (in pixels)	Thickness range (μm)	Volume range (US gallons)
True color	5528	[50, 500]	$7.30 \times 10^9 - 7.30 \times 10^{10}$
True color	583	[50, 500]	$7.70 \times 10^8 - 7.70 \times 10^9$
Sheen	6329	[0.04, 0.3]	$8.35 \times 10^8 - 8.35 \times 10^9$
Sheen	21,060	[0.04, 0.3]	$2.78 \times 10^9 - 2.78 \times 10^{10}$
Rainbow	120,074	[0.3, 5.0]	$8.94 \times 10^9 - 8.94 \times 10^{10}$

Table 4 IPW ratios for different hardware accelerators

Hardware accelerator	IPW ratio
RockChip RK3588S NPU	4.09 IPW
Intel Movidius Myriad X	18.82 IPW

Hardware evaluation

To ensure the adaptability of our proposed OilSpillNets framework for real-time applications, the models' performance is evaluated on two hardware accelerators: the RockChip RK3588S NPU and the Intel Movidius Myriad X, focusing on their efficiency in executing neural network inference tasks. The Inference Per Watt (IPW) ratio serves as a critical metric for assessing energy efficiency, especially in a resource-constrained environment. The RockChip RK3588S NPU demonstrates an IPW ratio of approximately 4.09 IPW when HDMI is connected. This indicates a moderate level of energy efficiency, balancing performance and power consumption during inference tasks. In contrast, the IntelMovidius Myriad X shows a significantly higher IPW ratio of 18.82 IPW. This remarkable efficiency highlights its suitability for real-time applications that require low power consumption without sacrificing performance (Table 4).

The comparison highlights the Intel Movidius Myriad X's advantage in energy efficiency, making it a compelling choice for applications that require sustained inference over extended periods. The RockChip RK3588S NPU, while less efficient, still provides a viable option for real-time applications such as marine monitoring and alerts/warnings.

Discussion

While SAR imagery remains a cornerstone for oil spill detection due to its all-weather, day-and-night capabilities, its inability to capture fine-scale properties prompted our integration of RGB drone data. By fusing SAR segmentation with high-resolution drone-based thickness estimates, we demonstrated a framework capable of both accurate spill delineation and preliminary property analysis. Our SAR segmentation model achieved an average IoU of 94.07%, and the RGB drone instance segmentation model

reached a Box (mAP50) of 79.4% and a Mask (mAP50) of 75.8%, underscoring the effectiveness of each modality in its respective domain. These results indicate that SAR excels at providing reliable two-dimensional spill footprints, while drone imagery contributes essential information about oil thickness, together enabling more comprehensive monitoring than either modality alone.

Limitations

Despite these promising results, several limitations constrain the broader applicability and robustness of our approach. Data generalisation remains a concern, as both our SAR and RGB drone datasets are geographically confined to the northern Mediterranean coast of Egypt, and their modest sizes, 691 SAR patches and 148 unique drone images, may not capture the full variability of oil types, sea states, or lighting conditions encountered in other regions. Consequently, model performance could degrade when applied to different coastlines or during severe weather events. Fusion synchronisation issues also arise because SAR acquisitions and drone flights are not perfectly simultaneous; temporal mismatches between SAR passes and drone overflights can cause the fused footprint to misrepresent a rapidly changing spill. Geospatial alignment challenges further complicate matters: projecting high-resolution drone pixels onto coarser SAR grids can introduce errors, since small georeferencing inaccuracies inherent to both SAR and drone platforms may lead to misaligned masks and incorrect volume estimates. Sensor calibration and radiometric consistency remain potential obstacles as well. Variations in SAR sensor parameters (wavelength, polarisation, incidence angle) and drone camera settings (exposure, white balance) affect the appearance of oil slicks and thickness cues. Lastly, our assumption of uniform thickness within each categorical range ("Rainbow," "Sheen," "True colour") introduces uncertainty into volume estimates, since actual oil thickness can vary significantly within these ranges. Consequently, our reported volume ranges represent bounding estimates rather than precise measurements.

Conclusion

This study demonstrates the technical benefits of a multi-modal data fusion approach for oil spill detection and property estimation. Specifically, we combined SAR Sentinel-1 imagery with high-resolution RGB drone imagery to detect and segment oil spill regions. The segmentation results were then used to estimate key physical characteristics of the spill, notably thickness, which enabled the calculation of oil slick volume.

To ensure the practicality of our method, we deployed and benchmarked the trained deep learning models on two edge AI platforms: the Intel Movidius MyriadX and the Rockchip RK3588S. We achieved a promising trade-off between

segmentation accuracy, inference latency, and power consumption. In particular, the RK3588S demonstrated strong performance for real-time processing of high-resolution input data, while maintaining efficiency suitable for field deployment. Our analysis confirms that integrating visual data from RGB and SAR sources enhances detection robustness and allows more accurate physical characterisation of oil spills under varying environmental conditions.

Appendix A: Additional qualitative examples

See Figs. 4, 5 and 6.

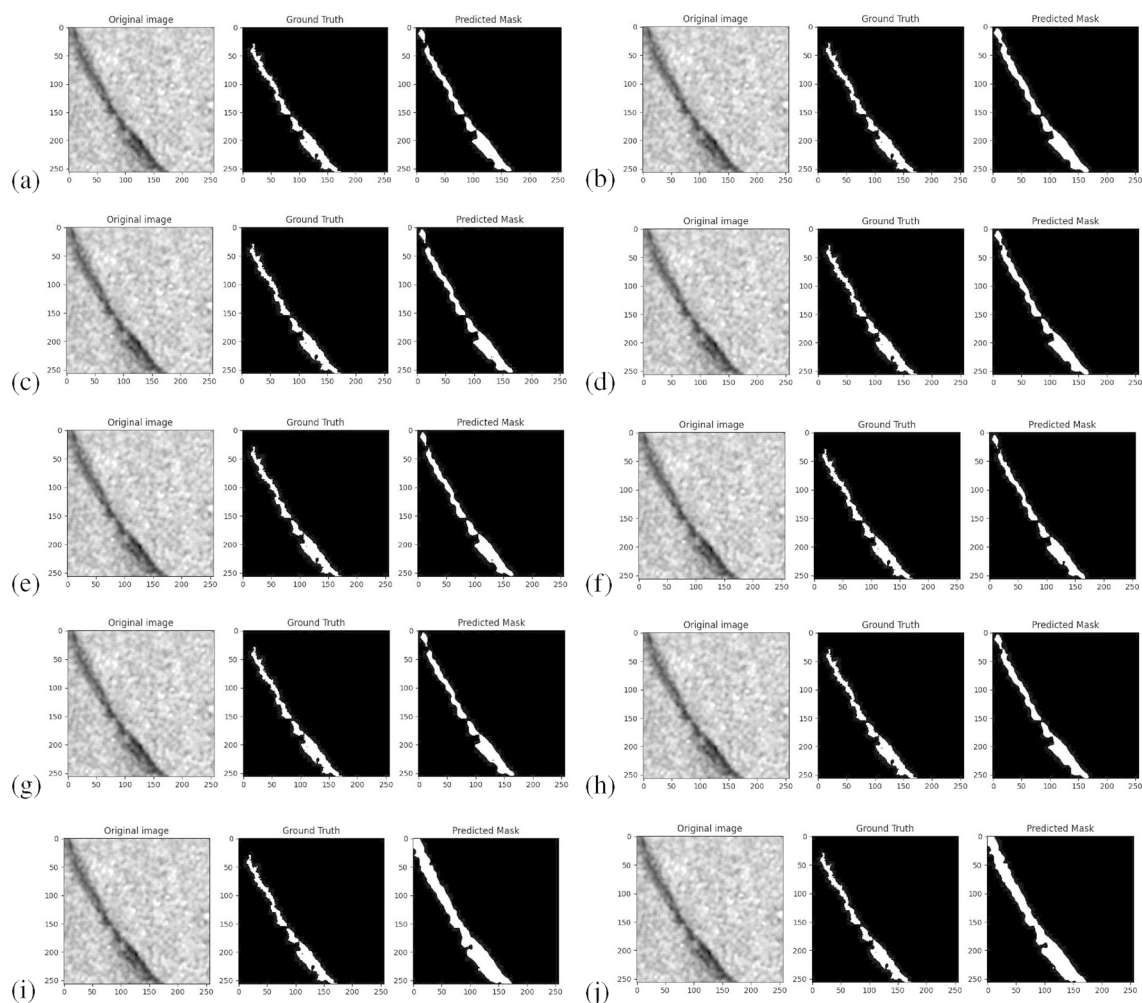


Fig. 4 Qualitative segmentation on test sample #1 (low-contrast conditions) for each backbone and training mode: **a** DeepLabV3-MobileNetV3 Large, frozen backbone; **b** DeepLabV3-MobileNetV3 Large, fine-tuned; **c** DeepLabV3-ResNet101, frozen; **d** Deep-

LabV3-ResNet101, fine-tuned; **e** DeepLabV3-ResNet50, frozen; **f** DeepLabV3-ResNet50, fine-tuned; **g** FCN-ResNet101, frozen; **h** FCN-ResNet101, fine-tuned; **i** FCN-ResNet50, fine-tuned; **j** FCN-ResNet50, frozen

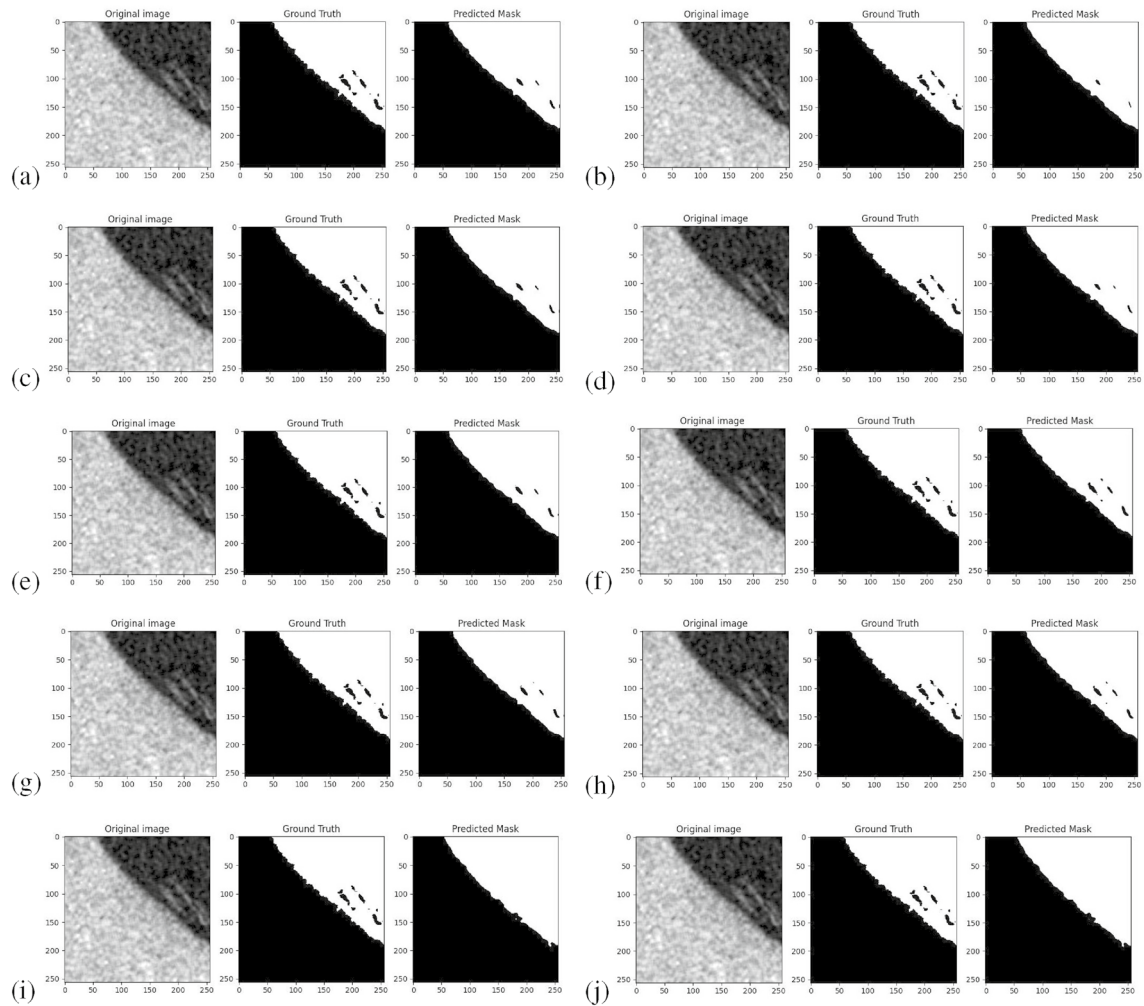


Fig. 5 Qualitative segmentation on test sample #2 (complex shoreline geometry) for each backbone and training mode: **a** DeepLabV3-MobileNetV3 Large, frozen backbone; **b** DeepLabV3-MobileNetV3 Large, fine-tuned; **c** DeepLabV3-ResNet101, frozen; **d** Deep-

LabV3-ResNet101, fine-tuned; **e** DeepLabV3-ResNet50, frozen; **f** DeepLabV3-ResNet50, fine-tuned; **g** FCN-ResNet101, frozen; **h** FCN-ResNet101, fine-tuned; **i** FCN-ResNet50, fine-tuned; **j** FCN-ResNet50, frozen

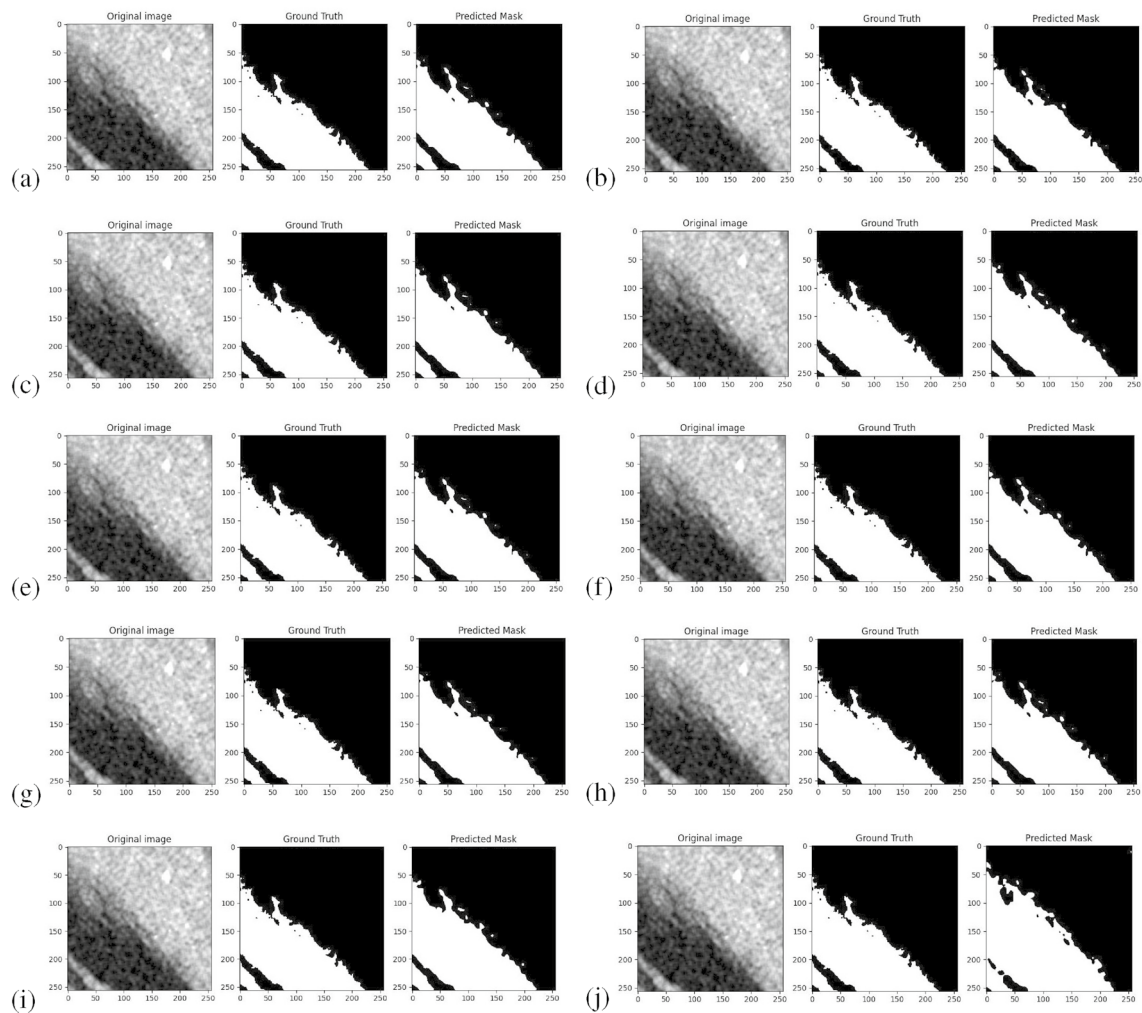


Fig. 6 Qualitative segmentation on test sample #3 (varying slick thickness) for each backbone and training mode: **a** DeepLabV3-MobileNetV3 Large, frozen backbone; **b** DeepLabV3-MobileNetV3 Large, fine-tuned; **c** DeepLabV3-ResNet101, frozen; **d** Deep-

LabV3-ResNet101, fine-tuned; **e** DeepLabV3-ResNet50, frozen; **f** DeepLabV3-ResNet50, fine-tuned; **g** FCN-ResNet101, frozen; **h** FCN-ResNet101, fine-tuned; **i** FCN-ResNet50, fine-tuned; **j** FCN-ResNet50, frozen

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability We have cited the data used, which is publicly available.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahmed S, ElGharbawi T, Salah M, El-Mewafi M (2023) Deep neural network for oil spill detection using Sentinel-1 data: Application to Egyptian coastal regions. *Geomat Nat Haz Risk* 14(1):76–94

- Al-Ruzouq R, Gibril M, Shanableh A, Kais A, Hamed O, Al-Mansoori S, Khalil MA (2020) Sensors, features, and machine learning for oil spill detection and monitoring: A review. *Remote Sensing* 12(20):3338
- Beyer J, Trannum HC, Bakke T, Hodson PV, Collier TK (2016) Environmental effects of the Deepwater Horizon oil spill: A review. *Mar Pollut Bull* 110(1):28–51
- Bianchi FM, Espeseth MM, Borch N (2020) Large-scale detection and categorisation of oil spills from SAR images with deep learning. *Remote Sensing* 12(14):2260
- Bonn Agreement (2016) *Special on volume calculation*. Microsoft Word document, . Available online: https://www.bonnagreem.org/site/assets/files/1081/special_on_volume_calculation_20160607.docx (accessed 7 June 2016)
- Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 40(4):834–848

- Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818
- Ciaglia F, Zuppichini FS, Guerrie P, McQuade M, Solawetz J (2022) Roboflow 100: A rich, multi-domain object detection benchmark. *arXiv*, [arXiv:2211.13523](https://arxiv.org/abs/2211.13523)
- De Kerf T, Gladines J, Sels S, Vanlanduit S (2020) Oil spill detection using machine learning and infrared images. *Remote Sensing* 12(24):4090
- De Kerf T, Sels S, Samsonova S, Vanlanduit S (2024) Oil spill drone: A dataset of drone-captured, segmented RGB images for oil spill detection in port environments. *arXiv*, [arXiv:2402.18202](https://arxiv.org/abs/2402.18202)
- Fingas M (2021) Visual appearance of oil on the sea. *Journal of Marine Science and Engineering* 9(1):97
- International Tanker Owners Pollution Federation (ITOPF). *Oil tanker spill statistics 2019*. Montreal, QC, Canada, (2019). Available online: https://www.itopf.org/fileadmin/data/Documents/Company_Lit/Oil_Spill_Stats_brochure_2020_for_web.pdf (accessed 14 December 2020)
- Janati M, Kolahdoozan M, Imanian H (2020) Artificial neural network modeling for the management of oil slick transport in marine environments. *Pollution* 6(2):399–415
- Krestenitis M, Orfanidis G, Ioannidis K, Avgerinakis K, Vrochidis S, Kompatsiaris I (2019) Oil spill identification from satellite images using deep neural networks. *Remote Sensing* 11(15):1762
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440
- Lynn T (2023) Detect and segment oil spills using computer vision. *Roboflow Blog*, 25 July. Available online: <https://blog.roboflow.com/detect-analyze-oil-spills-computer-vision/> (accessed 25 July 2023)
- Mahmoud AS, Mohamed SA, El-Khoriby RA, AbdelSalam HM, El-Khodary IA (2023) Oil spill identification based on dual attention UNet model using synthetic aperture radar images. *Journal of the Indian Society of Remote Sensing* 51(1):121–133
- Petrongonas E et al (2021) Paralos: A scheduling and memory management framework for heterogeneous VPUs. In *24th Euromicro Conference on Digital System Design (DSD)*, IEEE,
- Readman JW, Bartocci J, Tolosa I, Fowler SW, Oregioni B, Abdulraheem MY (1996) Recovery of the coastal marine environment in the Gulf following the 1991 war-related oil spills. *Mar Pollut Bull* 32(6):493–498
- Reis D, Kupec J, Hong J, Daoudi A (2023) Real-time flying object detection with YOLOv8. *arXiv*, [arXiv:2305.09972](https://arxiv.org/abs/2305.09972)
- Rockchip (2021) *Rockchip RK3588S Datasheet*. Available online: <https://www.cnx-software.com/pdf/datasheet/Rockchip-RK3588S-Datasheet%20V1.0-20211221.pdf>
- Statista. *Largest global oil spills since 1967*. Available online: <https://www.statista.com/statistics/268148/largest-global-oil-spills-since-1967/> (accessed 25 August 2023)
- Yan Z, Chong J, Zhao Y, Sun K, Wang Y, Li Y (2019) Multifeature fusion neural network for oceanic phenomena detection in SAR images. *Sensors* 20(1):210

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.