Deep Learning-Based Hyperspectral Oil Spill Detection for Marine Pollution Monitoring in the Gulf of Mexico: A Step Marine Pollution Monitoring and SDG 14 Compliance

Mahsa Samkhaniani[a]*, Afshin Khoshand[a] , Soran Ezati[b]

[a] Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran

[b] Environmental Engineering Division, Civil & Environmental Engineering Faculty, Tarbiat Modares University

* Corresponding author; E-mail address: m.samkhanian@kntu.ac.ir

Abstract

Oil spills pose a significant threat to marine ecosystems, undermining global efforts toward sustainable development goals (SDGs), particularly in preserving marine resources, oceans, and waters. This study investigates the application of hyperspectral remote sensing data combined with deep learning techniques to enhance oil spill detection accuracy and efficiency. Using the publicly available Hyperspectral Oil Spill Database (HOSD), we applied dimensionality reduction via Principal Component Analysis (PCA) and standardized patch-based input preparation. Two widely used convolutional neural network architectures, U-Net and DeepLabv3, were employed for semantic segmentation of oil-contaminated regions. Performance was assessed using standard metrics, including Intersection over Union (IoU) and F1-score. DeepLabv3 achieved higher precision and a slightly better IoU (57.86%) compared to U-Net. Given the inherent class imbalance in the data, IoU is considered a more appropriate evaluation metric, as it effectively balances false positives and false negatives during segmentation assessment. These promising results underscore the potential of hyperspectral imaging combined with deep learning for effective oil spill detection, thereby contributing to enhanced environmental monitoring in alignment with the sustainable development goals.

Keywords: Oil spill detection, Hyperspectral imaging, Marine pollution, Deep learning, SDG 14

Highlights

- DeepLabV3 achieved superior performance, with higher IoU (57.86%) and F1-score (73.3%) than U-Net.

- Oil spill detection using hyperspectral data demonstrated strong results despite differences from SAR-based studies.
- PCA-based dimensionality reduction improved computational efficiency while preserving key spectral features.
- U-Net outperformed in sensitivity, while DeepLabV3 showed higher precision, demonstrating their complementary strengths.

1. Introduction

Oil spills in the Gulf of Mexico continue to pose significant risks to marine ecosystems and wildlife. The Deepwater Horizon (DWH) oil exploration platform's explosion on April 20, 2010, resulted in a catastrophic release of approximately 640 million liters of crude oil into the northern Gulf, creating an oil slick that covered more than 112,000 km2 of the ocean's surface(Beyer et al., 2016). Over the past five decades, the number of acute oil spills (exceeding 7 tons) from vessels has shown a marked decline. The global count dropped from 94 incidents in the 1980s to 77 in the 1990s, 32 in the 2000s, 18 in the 2010s, and only 5 reported cases so far in the current decade .This substantial reduction occurred despite a near doubling in the volume of crude oil and other tanker trade during the same period. Nonetheless, oil spills continue to occur, and maintaining robust contingency plans and rapid-response capabilities remains essential to mitigating their environmental impact(Brandvik et al., 2025). Even small spills can have profound and long-lasting effects. While the immediate impacts may diminish, it often takes years or even decades for an affected area or ecosystem to fully recover. The recovery timeline depends on various factors, including the spill's characteristics and the specific conditions of the impacted area (Prasad & Anuprakash, 2016). The toxicity of oil spills stems from the presence of various hazardous compounds, such as Polycyclic Aromatic Hydrocarbons and volatile organic compounds (Massoura & Sommerville, 2009). Upon release, oil typically forms thick layers on the water's surface, obstructing essential processes such as oxygen exchange and severely impacting marine life, including their nervous systems (Jha & Dahiya, 2022). The impact of oil spills on the marine environment, ecosystems, and society is multifaceted. It includes effects such as altered flow rates, the occurrence of poisonings, DNA damage, loss of buoyancy, and economic repercussions These effects can have both short- and long-term consequences, influencing employment, public health, and social welfare (Chang et al. 2014). While each oil spill is distinct, existing literature identifies several critical variables that influence the frequency and severity of such incidents, including the spill's volume, physical extent, and the effectiveness of disaster management response (Chang et al. 2014). According to Goal 14 of the SDGs, which focuses on the conservation and sustainable use of oceans, seas, and marine resources, monitoring and mitigating the effects of oil spills is crucial for achieving

this goal (Sivadas et al., 2021). Therefore, monitoring and supervision of pollution caused by oil spills to prevent its adverse effects plays an important role to achieve this goal. Remote sensing techniques enable the monitoring of oil spill spatial distribution in remote and inaccessible large areas (Bing et al., 2019). The synergistic combination of DL models and Earth observation holds significant promise for advancing the support of SDGs (Persello et al., 2022). Today, various platforms such as vessels, aircraft, and satellites are employed for oil spill monitoring, with remote sensing becoming a key tool due to its rapid, wide-area coverage and ability to provide vital information like oil type and thickness(Brekke & Solberg, 2005). Among these, SAR and optical sensors—particularly multispectral and hyperspectral—are the most widely used approaches for large-scale marine oil spill detection(Xie et al., 2025). This variety of sensors and their capabilities gives an advantage to analysts to interpret more than what a human naked eye can detect (a human eye is able to detect only the visible portion of the electromagnetic spectrum (Brekke & Solberg, 2005; Rajendran et al., 2021). Active microwave sensors like Synthetic Aperture Radar (SAR) captures two-dimensional images. The image brightness is a reflection of the microwave backscattering properties of the surface. SAR deployed on satellites is today an important tool in oil spill monitoring due to its wide area coverage and day and night all-weather capabilities (Raju et al., 2024). A major challenge in detecting oil spills with SAR data is differentiating oil spill regions from other natural phenomena. Dark patches in SAR images could be caused by factors such as grease ice, wind sheltering by land, or internal waves (Ajadi et al., 2018). Moreover, SAR data has other limitations, including high costs, low revisit frequencies, and relatively narrow swath widths. These constraints necessitate the advancement of multi-platform SAR systems and the use of complementary optical sensors, such as multispectral and hyperspectral sensors (Duan et al., 2023). Hyperspectral remote sensing is an excellent tool for detecting and classifying oil slicks due to its rich spectral information (Gou et al., 2025). Multispectral data, with fewer bands than hyperspectral data, may miss valuable spectral details needed for identifying and classifying oil films. In contrast, hyperspectral remote sensing provides nearly continuous spectral data for both oil films and water. This continuous and detailed spectral information makes hyperspectral remote sensing particularly effective for oil spill detection. It offers higher spectral resolution than broadband spectral remote sensing, allowing for the identification of subtle but important features (Liu et al., 2018). To achieve more accurate classification, spectra extracted from hyperspectral images were more reliable due to same observation conditions (Liu et al., 2016). Hyperspectral images provide a high spectral resolution, allowing for the detection of subtle variations in the reflectance of different materials, including oil spills (Zhao et al., 2022). Hyperspectral sensors mounted on aircraft represent an advanced technique in remote sensing, offering both high spectral and spatial resolution. These sensors can capture hundreds of narrow spectral channels, covering wavelengths from the visible to the infrared spectrum (Ghamisi et al., 2018). Hyperspectral imaging has opened up a wide range of applications across various Earth sciences (Goetz, 2009).

Environmental studies utilizing hyperspectral imaging are producing results that would be unattainable or too costly and time-consuming with conventional methods. A notable example is the cleanup of the Leadville, CO Superfund Site, where AVIRIS imagery, combined with field spectral measurements, identified waste piles with the highest risk of heavy metal leaching into streams and groundwater—saving millions of dollars in remediation efforts (Goetz, 2009). Oil spills present a significant threat to marine ecosystems, necessitating rapid and accurate detection methods (Huby et al., 2022). Traditional approaches relying on optical and SAR imagery often suffer from inaccuracies due to confounding factors like grease ice (Vijayakumar, 2023). DL has emerged as a powerful tool for enhancing oil spill detection in hyperspectral data. Leveraging Convolutional Neural Networks (CNNs), inspired by the human brain, DL excels at extracting intricate patterns from vast datasets. Unlike conventional machine learning techniques, DL autonomously learns features directly from the data, making it well-suited for analyzing hyperspectral imagery, which captures detailed spectral information across numerous narrow bands (Huby et al., 2022).Various DL strategies have been employed for oil spill detection. These include image classification, pixel-based classification, object detection, semantic segmentation, and instance segmentation. Each method offers distinct advantages in accurately identifying and delineating oil spills (Prasad et al., 2022).Several DL algorithms have been utilized for oil spill detection, including U-Net, FC-DenseNet, DeepLabV3+ (Dehghani-Dehcheshmeh et al., 2023), U-Net with EfficientNet-B3 backbone (de Moura et al., 2022) , Faster R-CNN (Huang et al., 2022), Mask R-CNN(Yekeen et al., 2020), DenseNet201 combined with U-Net (Huby et al., 2022), VGG16, Mask R-CNN, PSPNet(Ghorbani & Behzadan, 2021) and others. Each algorithm exhibits unique advantages and performance metrics in detecting and segmenting oil spills. Several studies have explored semantic segmentation using deep convolutional neural networks (DCNNs) for oil spill detection in SAR imagery. These studies have demonstrated the effectiveness of DCNN architectures like U-Net and DeepLabV3+ in achieving acceptable results for oil spill segmentation in SAR dataSaeed Dehghani et al. (2023) reached the F1-score values of 61.58% for the U-Net algorithm and 75.8% for the DeepLabV3+ algorithm for detecting oil spills(Dehghani-Dehcheshmeh et al., 2023). Krestenitis et al. compared the performance of several architectures, including U-Net, LinkNet, PSPNet, DeepLabv2, and DeepLabv3+. The results showed that the DeepLabv3+ model achieved the best performance, with a mean Intersection over Union (mIoU) of 65%. The U-Net algorithm, achieved a lower mIoU of 53.38% compared to DeepLabv3+(Krestenitis et al., 2019).In 2023, Maruf A. Dhali et al test the feasibility of deep encoder-decoder models that can be effectively trained for oil spill detection. The best model is the one with ResNet-50 encoder and DeepLabV3+ decoder. The IoU values obtained for the oil class in this study were equal to 53.38%(Satyanarayana & Dhali, 2023). In 2022, Najla Vilar Aires de Moura and her colleagues investigated three DL models (U-net, DeepLabv3+, and LinkNet) to monitor oil spills in the Brazilian territory. The dataset was manually labeled with confirmation from the oil company and federal inspection,

resulting in *800* images with *512 × 512* spatial dimensions and their respective annotations (600 for training, 125 for validation, and 75 for testing). They achieved IoU values ranging from 56% to 61% for DeepLabV3 and 55% to 62% for U-Net, with F-score values approximately between 71% and 76% across three backbones (de Moura et al., 2022). Unlike earlier studies that primarily relied on Synthetic Aperture Radar (SAR) data, this research utilizes the publicly available Hyperspectral Oil Spill Database (HOSD). Hyperspectral imagery provides significantly richer spectral information than SAR, enabling more accurate differentiation between oil-contaminated regions and the surrounding water. This enhanced spectral resolution allows deep learning models to detect subtle spectral variations, leading to improved segmentation performance. By assessing the performance of U-Net and DeepLabv3 on hyperspectral data, this study presents a scalable and adaptable alternative for oil spill detection. The proposed approach contributes to more reliable marine environmental monitoring and aligns with the objectives of Sustainable Development Goal 14, which advocates for the conservation and sustainable use of oceans and marine resources.To improve computational efficiency and focus on the most informative spectral features, a series of preprocessing steps—including Principal Component Analysis (PCA)—were applied to reduce the dimensionality of the hyperspectral data. The images were then divided into standardized patches to augment the dataset and ensure compatibility with the input requirements of the U-Net and DeepLabv3 architectures. These models were subsequently trained and evaluated on the processed data, providing insights into their effectiveness for hyperspectral oil spill segmentation.

2. Dataset description and pre-processing

2.1. Hyperspectral Dataset for Oil Spill Detection in the Gulf of Mexico

The Deepwater Horizon oil spill, the largest marine oil spill in U.S. history, occurred on April 20, 2010, in the Gulf of Mexico near 25°N, 90°W (Fig. 1). Accurate detection of the affected areas is crucial for effective monitoring and cleanup, and hyperspectral data, covering the visible to infrared spectrum, is a key tool for this purpose . The HOSD, the first public dataset for oil spill detection, supports the development of advanced detection methods. Built using AVIRIS sensor data, it offers wide coverage and the largest dataset available. Oil spill regions in the database are manually annotated pixel by pixel in ENVI software under expert guidance. The dataset covers a spectral range from 365 nm to 2500 nm, with spatial resolution varying based on flight altitude (Duan et al., 2023).

Fig. 1. Location of the Deepwater Horizon oil spill in the Gulf of Mexico

2.2. Image pre-processing

Hyperspectral imaging through remote sensing is a highly dynamic area of research, with numerous contributions in the recent scientific literature. Analyzing these images poses significant computational challenges, primarily due to their high-dimensional nature and the complexity of state-of-the-art algorithms designed for hyperspectral image processing (Qian, 2022).While hyperspectral images offer rich information about land objects, they often contain hundreds of spectral bands with noise and redundancy. These characteristics can hinder classification and analysis (Fernandez et al., 2019)

Here, pre-processing techniques like dimensionality reduction come into play. Dimensionality reduction aims to identify the most informative subset of features, removing redundancies while preserving key spectral details (Ghamisi et al., 2017). Dimensionality reduction can be achieved through two main approaches: (1) feature selection and (2) feature extraction. Reducing the number of bands improves computational efficiency because processing smaller data sets significantly reduces computational demands, making analysis faster, especially for real-time applications such as fire detection or environmental monitoring (Ustuner, 2024).

It also reduces the need for data storage space. Smaller data sets require less storage space, which makes data management more manageable. Therefore, preprocessing hyperspectral images through dimensionality reduction is very important not only to reduce noise and redundancy, but also to enable more efficient analysis and data management (Fernandez et al., 2019)

2.2.1. PCA

Hyperspectral images typically consist of hundreds of spectral bands, many of which exhibit significant redundancy due to high inter-band correlations (Licciardi, 2019). This redundancy increases computational complexity and may introduce noise, negatively impacting classification performance(Licciardi, 2019). PCA is a widely adopted dimensionality reduction technique that projects high-dimensional data into a lower-dimensional space while preserving the most significant spectral variations (Hossain et al., 2012). PCA operates by computing a new set of orthogonal components, ranked by their ability to capture variance within the dataset. The first few components retain the maximum amount of meaningful spectral information, while the remaining components, which primarily capture noise and redundant variations, are discarded (Shabna & Ganesan, 2014). By transforming the original spectral space into an optimized feature subspace, PCA improves computational efficiency and enhances the robustness of machine learning and DL models applied to hyperspectral data analysis (Uddin et al., 2019).In this study, the original 224 spectral bands from AVIRIS hyperspectral imagery were reduced to 32 principal components, capturing the most discriminative spectral features. This reduction not only decreases the computational burden but also mitigates the risk of overfitting in DL models by eliminating redundant information. The transformed feature space provides a compact and optimized representation of the data, ultimately improving segmentation accuracy for oil spill detection .

2.2.2. Standardizing Input and Enhancing Efficiency

The HOSD image dataset consists of 18 images with varying dimensions, as summarized in the table below.

Table 1. Overview of HOSD Image Dataset: Spatial Size, Flight Time, and Resolution

In order to ensure standardized input for the DL algorithm, pre-processing involves transforming these images into *256×256* patches (Fig. 2). Given the diverse nature of hyperspectral data, characterized by differing dimensions and spectral properties, a critical pre-processing stage is essential to normalize the dataset for comprehensive analysis. To ensure uniformity and streamline algorithmic processing, each image is resized into *256×256*pixel patches. This standardization guarantees consistent input dimensions for subsequent analysis. The choice of patch size depends on various factors, including the specific application, the characteristics of the images, and the desired level of accuracy.

The patch sizes used in these previous studies varied. Some studies used smaller patch sizes, such as 28 × 28 pixels, 50 × 50 pixels, 128 × 128 pixels, 160 × 160 pixels, 256 × 256 pixels, 320 × 320 pixels, 321 × 321 pixels, and 336 × 336 pixels(Cantorna et al., 2019; Krestenitis et al., 2019; Li et al., 2021; Nieto-Hidalgo et al., 2018; Wang et al., 2021).Previous studies have shown that larger window sizes can minimize

the interference caused by errors at patch edges. In contrast, smaller patch sizes may concentrate errors and lead to less accurate segmentation results(Huang et al., 2022; Huang et al., 2023)

Furthermore, to enhance the efficiency of the oil spill detection algorithm, a targeted filtering process is applied to the dataset. Images featuring oil spill areas are retained for analysis, while those displaying complete darkness are excluded. This meticulous curation ensures that only pertinent data relevant to the detection task is utilized, thereby boosting accuracy and efficiency in subsequent analyses. Through these meticulous pre-processing steps, our objective is to prepare a refined dataset conducive to robust oil spill detection, thereby facilitating effective environmental monitoring and management.

Fig. 2. Three representative examples of the 256 × 256 pixel original images and their corresponding ground truth masks from the HOSD dataset.

3. Methodology

This study employs deep learning-based semantic segmentation to detect oil spill areas in hyperspectral data from the HOSD. After preprocessing the data using PCA and patch extraction, two segmentation models—U-Net and DeepLabv3—were trained and compared.

3.1. Deep learning models

For the autonomous oil spill detection system, two prominent deep learning architectures, U-Net and DeepLabv3, were utilized due to their recognized proficiency in image segmentation tasks. Both models excel at classifying each pixel in an image, enabling precise localization of oil spills by distinguishing them from the background.

The U-Net architecture, renowned for picture segmentation, was employed for the autonomous oil spill detection system. U-Net features contracting and expanding routes for segmentation. Initially, the Contracting Path (Encoder) extracts relevant picture information meticulously. This involves a series of convolutional layers and pooling procedures, such as max pooling, to reduce picture spatial dimensions while increasing feature channels. This controlled resolution reduction captures progressively complex picture data characteristics. The Bridge segment connects the contracting and expanding routes, processing encoder-extracted high-level features using convolutional layers to prepare for decoding. In the Expanding Path (Decoder), the network restores spatial resolution lost during contracting, utilizing transposed convolutions for up-sampling to create a segmentation mask that matches the picture(Ronneberger et al., 2015). U-Net's effectiveness is partly attributed to its incorporation of skip connections, which directly copy feature maps from corresponding levels in the contracting path to the expanding path. This allows the decoder to access precise location information captured at earlier stages in the encoder, leading to more accurate segmentation results(Chen & Wang, 2022). As depicted in Fig. 3, which illustrates the utilized U-

Net architecture, images with dimensions of 256 (height) × 256 (width) × 32 (channels) serve as input. Features are extracted via two 3×3 layers within custom convolutional blocks, with optional batch normalization and ReLU activation applied for optimization and non-linearity. The concatenation of outputs from both convolutional layers within each block enriches feature learning. The final layer employs a (1,1) convolution to map features to desired classes, followed by a Softmax activation for probability distributions.

Complementing U-Net,DeepLabv3, a convolutional neural network (CNN) architecture, is renowned for its proficiency in oil spill detection through semantic segmentation(Chen et al., 2017). In DeepLabv3, each pixel in an image is classified as either "oil spill" or "background," enabling precise spill localization, akin to U-Net. Typically, DeepLab employs an encoder-decoder structure similar to U-Net, where the encoder extracts informative features from the input image, and the decoder refines these features to produce a segmentation map with the same dimensions as the original image (Shanmukh et al., 2024). This segmentation map assigns a class (oil spill or background) to each pixel, facilitating visualization of the detected spill area. A powerful technique utilized by DeepLabv3 is atrous convolution (dilated convolution), enabling the network to capture long-range dependencies within the image without compromising resolution. This capability is crucial for accurately detecting oil spills, which often span large areas and necessitate consideration of the broader context, such as surrounding water(Chen et al., 2017). The DeepLabv3 model architecture used in this study, as shown in Fig. 4, comprises an encoding path, a decoding path, and a unique Atrous Spatial Pyramid Pooling (ASPP) module positioned between them. The encoding path focuses on feature extraction and dimension reduction via custom convolutional blocks with increasing filter sizes and max-pooling operations. The ASPP module is crucial, containing four parallel convolutional blocks with distinct dilation rates (6, 12, and 18), enabling the model to capture multi-scale features essential for accurate segmentation. The decoding path mirrors the encoding path, incorporating corresponding feature maps from the encoder to preserve spatial details lost during down-sampling. The final output layer uses a (1,1) convolution to map features to the desired classes, followed by a Softmax activation function for probability distributions.

(a)

(b)

Fig. 4. Architectures of deep learning models for oil spill detection using hyperspectral imagery: (a) U-Net model, featuring a contracting path for feature extraction, an expanding path with skip connections for precise segmentation, and 256×256×32 input dimensions; (b) DeepLabv3 model, incorporating an encoding path, Atrous Spatial Pyramid Pooling (ASPP) with dilation rates of 6, 12, and 18, and a decoding path for accurate segmentation of 256×256×32 hyperspectral images.

3.2. Model configurations

Two segmentation models are explored, U-Net and DeepLabv3. U-Net has a smaller number of parameters compared to DeepLabv3, making it potentially faster to train. The number of parameters is directly influenced by the model architecture (Park et al., 2023). A more complex architecture like DeepLabv3 with its ASPP module will generally have a higher number of parameters compared to U-Net. The number of parameters of each model is given in Table 2.

Table 2. Comparison of model size and parameter count for U-Net and DeepLabv3

Both models were optimized using the Adam optimizer, selected for its efficient convergence in deep networks by adaptively adjusting learning rates based on first and second-order moments (Kinga & Adam, 2015). An initial learning rate of 0.001 was set to balance convergence speed and stability (Loshchilov & Hutter, 2016), and training was conducted for a maximum of 200 epochs. Categorical cross-entropy was employed as the loss function to effectively optimize pixel-wise classification in multi-class semantic segmentation tasks (LeCun et al., 2015). To prevent overfitting, early stopping based on validation loss was applied, terminating training when no further improvement was observed (Prechelt, 2002).

3.3. Model validation

outputs of the models with the Softmax activation function had two bands corresponding to the oil spill and sea water classes. From the confusion matrix, various metrics are derived, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. These measures define the following metrics for assessing the models' performance: Accuracy, Recall, Precision, F1-score, and IoU.F1-score, a metric combining precision and recall, is valuable for evaluating the performance of semantic segmentation models. In semantic segmentation, precisely classifying each pixel (high precision) can be at odds with correctly identifying all pixels of a particular class (high recall). The F1-score balances these competing factors into a single score(Cui et al., 2023). IoU is the most stringent metric, typically yielding lower values compared to others. It measures the overlap between the predicted and ground truth

images, penalizing both false positives and false negatives(Rahman & Wang, 2016). These five-evaluation metrics are defined in Eq. (1) (2) (3) (4) (5).

$$\text{Accuracy} = \frac{(TP + TN)}{TP + FN + TN + FP} \quad (1)$$

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

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{F1} - \text{score} = \frac{2(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (4)$$

$$\text{IoU} = \frac{TP}{(TP + FP + FN)} \quad (5)$$

4. Results and discussions

4.1. Data Processing and Model Implementation

The acquisition of a noise-free, unified, and optimized dataset containing the most pertinent spectral information was facilitated through the application of PCA on the dataset. The most relevant spectral features were extracted while minimizing the number of required bands. Subsequent dimensionality reduction significantly accelerated the model's processing speed and reduced memory consumption. Furthermore, noisy and irrelevant spectral information was effectively filtered out by PCA, leading to a more robust and reliable analysis compared to traditional methods. A balance between retaining vital spectral information and achieving computational efficiency was ensured through the selection of the chosen components. A comprehensive evaluation was conducted to identify the most suitable number for this dataset, revealing that a slightly larger set of components, exceeding 25, yielded superior performance in capturing the key spectral information relevant to the investigation, building upon previous research that utilized 25 components(Duan et al., 2023).

Data augmentation was employed to enrich the training data, given the limited size and dimensional inconsistencies of the initial dataset. The data architecture was restructured by segmenting the data points into smaller, uniformly sized patches of 256×256 pixels. This technique effectively increased the size and diversity of the training set, mitigating the limitations of the original data.

Data augmentation via image patching was employed to expand the initial dataset to 142 images with a consistent size of 256×256 pixels. To ensure robust model evaluation, this augmented dataset was subsequently divided into training and testing sets. Following common practice, 70% (99 images) of the data were allocated for training the model, and the remaining 30% (43 images) were designated for testing its generalizability and performance.

A combination of U-Net and DeepLabv3 architectures was deliberately chosen for implementation in the image segmentation analysis. Table 2 displays the total number of training parameters. In DL models, the network training speed and the number of parameters are among the influential factors. An increase in the number of parameters typically leads to a longer network training time(Park et al., 2023).

4.2. Model Performance Evaluation

Table 3 presents the metric results of the models. Among the evaluation metrics, the lowest value is attributed to the IoU. Oil spills typically exhibit a significant class imbalance, with the majority of pixels belonging to the background class(Bui et al., 2024). In such scenarios, F1-score, a metric combining precision and recall, can be a valuable alternative. F1-score balances the trade-off between precisely classifying each pixel (high precision) and capturing all relevant pixels within the oil spill (high recall) (Cui et al., 2023). Thus, the overall accuracy metric is very little informative, achieving more than 96% in both models, even with more significant differences in the other metrics.

Table 3. Comparison of segmentation models in terms of evaluation metrics (%)

A close competition between U-Net and DeepLabv3 architectures for image segmentation was revealed by the analysis. As shown in Table 3, DeepLabv3 achieved a slightly higher overall accuracy (97.04%) compared to U-Net (96.67%). However, both models demonstrated strengths in distinct areas. Oil spill datasets typically suffer from class imbalance. IoU, by evaluating the model's performance specifically on the oil spill class rather than overall pixel-wise accuracy, is robust to this imbalance. U-Net exhibited good sensitivity (63.34%), while DeepLabv3 displayed a stronger performance in this area (68.54%) (Table 3). This indicates superior ability of DeepLabv3 to identify true positive pixels (oil spill pixels) in the images. DeepLabv3 also achieved a significantly higher precision score (78.64%) compared to U-Net (76.56%) (Table 3), signifying fewer false positives. Additionally, a noticeably higher IoU score (57.86%) was achieved by DeepLabv3 compared to U-Net (53%). This suggests a better average overlap between the predicted and ground truth segmentation masks, potentially making DeepLabv3 more effective in scenarios where accurate target area capture is crucial. Despite the slight difference in F1-scores (U-Net: 69.32%, DeepLabv3: 73.30%), both models achieved a balanced performance between precision and recall. In conclusion, both U-Net and DeepLabv3 are powerful image segmentation models, but cater to different needs. U-Net prioritizes accurate true positive identification, making it ideal for minimizing false negatives. DeepLabv3 excels in reducing false positives and capturing the target area more accurately, making it suitable for scenarios where those factors are critical.

Fig. 5illustrates the variations of training and validation loss values over epochs for the DeepLabV3 oil spill detection model. The shape and dynamics of the loss curves are crucial for diagnosing the model's behavior (e.g., underfitting, overfitting, or well-fitting). Loss represents the cumulative difference between the model's predictions and the ground truth for each sample in the training or validation datasets. Lower loss values indicate predictions closer to the actual data (ground truth). As evident in Fig. 5, both training and validation loss curves exhibit a continuous decrease, eventually reaching stability with a small gap. This trend suggests a good fit, where the model learns without overfitting to the training data. Additionally, alongside the decreasing loss curves, we can expect a corresponding increase in validation accuracy, further confirming the model's effectiveness.

Promising results were yielded by analysis using F1-score and IoU for evaluating U-Net and DeepLabv3 architectures on oil spill detection with hyperspectral images (Table 3). An F1-score of 69.32% and an IoU of 53% were achieved by the U-Net model, while a F1-score of 73.30% and an IoU of 57.86% were achieved by DeepLabv3. Figure 6 illustrates the oil spill detection outputs of the DeepLabv3 model applied to hyperspectral imagery from the Hyperspectral Oil Spill Database (HOSD). The figure presents five representative examples of predicted segmentation maps, showcasing DeepLabv3's ability to delineate oil-contaminated regions from the background.

Fig. 5. Training and validation loss and accuracy for the DeepLabV3 Oil Spill Detection Model

Fig. 6. shows five examples of detected oil spills with Deeplabv3

4.3. Comparative Analysis with SAR-Based Oil Spill Detection

The results obtained in this study provide valuable insights when compared to recent studies on oil spill detection using SAR. For example, F1-scores of 61.58% for U-Net and 75.8% for DeepLabV3+, which shares a similar architecture with DeepLabV3, were reported by (Dehghani-Dehcheshmeh et al., 2023). Similarly, it was found by Krestenitis et al. that DeepLabV3+ outperformed U-Net, with a mean IoU of 65%, compared to 53.38% for U-Net(Krestenitis et al., 2019). Additionally, a model incorporating a ResNet-50 encoder and DeepLabV3+ decoder was tested by Maruf A. Dhali et al. with an IoU of 53.38% for the oil class being achieved(Satyanarayana & Dhali, 2023). However, direct comparison with SAR-based studies remains challenging due to the inherent differences in data modalities. While hyperspectral images used in this study offer rich spectral data, SAR imagery is characterized by all-weather and day-night operational capabilities, which can significantly influence model performance and detection accuracy(Duan et al., 2023). To address the high-dimensional nature of hyperspectral data, PCA-based

dimensionality reduction was applied, and data augmentation was employed to overcome the limitations caused by the dataset size. Moreover, alternative evaluation metrics (e.g., overall accuracy or probability of detection) are often reported in SAR-based studies, which complicates direct comparisons with the F1-score and IoU used in this study.

Despite these challenges, strong performance was demonstrated by the models. U-Net and DeepLabV3 performed well, with sensitivity being excelled by U-Net and higher precision being demonstrated by DeepLabV3. Future research could explore data fusion techniques, which would combine the complementary strengths of hyperspectral and SAR data to improve oil spill detection. Additionally, standardization of evaluation metrics across different data modalities is suggested to enable more meaningful comparisons between studies utilizing various remote sensing techniques.

5. Limitations and Future Work

This study yields promising results for oil spill detection using hyperspectral imagery and deep learning, yet limitations persist. The Hyperspectral Oil Spill Database (HOSD) lacks geographic and spectral diversity, potentially reducing model generalizability across diverse marine environments. Hyperspectral imaging also presents challenges, including high costs, large data volumes, and the need for specialized expertise. Future research should focus on integrating hyperspectral data with modalities like SAR or multispectral imagery, expanding datasets to include varied conditions, optimizing models for computational efficiency, and standardizing evaluation metrics to enhance comparability. These efforts will advance scalable, robust oil spill monitoring systems, supporting Sustainable Development Goal 14.

6.Conclusion

Oil spills remain a persistent threat to marine ecosystems, causing widespread ecological damage and jeopardizing wildlife populations. The catastrophic consequences of such events are exemplified by the Deepwater Horizon disaster. Despite ongoing efforts for prevention, occurrences of oil spills continue, underlining the critical need for improved detection and monitoring methods. The potential of DL models for oil spill detection in hyperspectral imagery was explored in this study. PCA was employed for dimensionality reduction, addressing the challenges associated with data scarcity and high dimensionality in real-world hyperspectral datasets. The dataset was expanded by segmenting images into standardized 256×256 pixel patches, increasing the training set size and diversity. The performance of U-Net and DeepLabv3 architectures was compared in our investigation. High overall accuracy was exhibited by both models; however, distinct strengths were demonstrated. Sensitivity was excelled in by U-Net, effectively identifying true positive oil spill pixels, while superior precision was demonstrated by DeepLabv3, minimizing false positives and achieving a higher IoU score. These findings suggest that the optimal choice between U-Net and DeepLabv3 depends on the specific requirements of the oil spill detection task. While a specific and limited dataset was focused on in this research, the way is paved for further exploration of DL models for oil spill detection using hyperspectral imagery. The effectiveness of these models on larger and more diverse datasets could be investigated in future studies, potentially incorporating data from different sensors or geographic regions. Ultimately, significant promise is held by the real-world implementation of DL models within oil spill detection systems using hyperspectral imagery. By harnessing advancements in remote sensing and artificial intelligence, contributions can be made to more effective environmental monitoring and protection of marine ecosystems, aligning with the essential goals of sustainable development.

Acknowledgment

The authors would like to sincerely thank Professor Pedram Ghamisi for his valuable guidance and unwavering support throughout the course of this research. His expert insights and constructive feedback greatly contributed to the methodological development and data interpretation. The authors are especially grateful for the access to the Hyperspectral Oil Spill Dataset (HOSD), which was crucial for model training and validation. His contributions significantly enhanced the overall quality and impact of this study.

Authors contribution statement All authors read and approved the final manuscript.

Mahsa Samkhaniani: Conceptualization, Methodology, Modeling, Data curation, Formal analysis, Visualization, Writing – original draft.

Afshin Khoshand : Conceptualization, Methodology, Formal analysis, review & editing, Supervision.

Soran Ezati : Conceptualization, Methodology, Writing – review & editing.

Data, Material and/or Code availability All Data, Material and/or Code available on request from the corresponding author.

References

Ajadi, O. A., Meyer, F. J., Tello, M., & Ruello, G. (2018). Oil spill detection in synthetic aperture radar images using Lipschitz-regularity and multiscale techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(7), 2389-2405.

Beyer, J., Trannum, H. C., Bakke, T., Hodson, P. V., & Collier, T. K. (2016). Environmental effects of the Deepwater Horizon oil spill: A review. *Marine pollution bulletin*, *110*(1), 28-51.

Bing, L., Xing, Q.-G., Liu, X., & Zou, N.-N. (2019). Spatial distribution characteristics of oil spills in the Bohai Sea based on satellite remote sensing and GIS. *Journal of Coastal Research*, *90*(SI), 164-170.

Brandvik, P. J., Faksness, L.-G., Daae, R. L., Skancke, J., Daling, P. S., & Øye, K. S. (2025). An evaluation of the climate effect of selected oil spill response technologies based on their aerial emissions. *Marine pollution bulletin*, *218*, 118169.

Brekke, C., & Solberg, A. H. (2005). Oil spill detection by satellite remote sensing. *Remote sensing of environment*, *95*(1), 1-13.

Bui, N. A., Oh, Y., & Lee, I. (2024). Oil spill detection and classification through deep learning and tailored data augmentation. *International Journal of Applied Earth Observation and Geoinformation*, *129*, 103845.

Cantorna, D., Dafonte, C., Iglesias, A., & Arcay, B. (2019). Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms. *Applied Soft Computing*, *84*, 105716.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.

Chen, Y., & Wang, Z. (2022). Marine oil spill detection from SAR images based on attention U-Net model using polarimetric and wind speed information. *International Journal of Environmental Research and Public Health*, *19*(19), 12315.

Cui, L., Wang, L., Su, J., Song, Z., & Li, X. (2023). Classification and identification of degraded alpine m eadows based on machine learning techniques. 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL),

de Moura, N. V. A., de Carvalho, O. L. F., Gomes, R. A. T., Guimarães, R. F., & de Carvalho Júnior, O. A. (2022). Deep-water oil-spill monitoring and recurrence analysis in the Brazilian territory using Sentinel-1 time series and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, *107*, 102695.

Dehghani-Dehcheshmeh, S., Akhoondzadeh, M., & Homayouni, S. (2023). Oil spills detection from SAR Earth observations based on a hybrid CNN transformer networks. *Marine pollution bulletin*, *190*, 114834.

Duan, P., Kang, X., Ghamisi, P., & Li, S. (2023). Hyperspectral remote sensing benchmark database for oil spill detection with an isolation forest-guided unsupervised detector. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-11.

Fernandez, D., Gonzalez, C., Mozos, D., & Lopez, S. (2019). FPGA implementation of the principal component analysis algorithm for dimensionality reduction of hyperspectral images. *Journal of Real-Time Image Processing*, *16*, 1395-1406.

Ghamisi, P., Maggiori, E., Li, S., Souza, R., Tarablaka, Y., Moser, G., De Giorgi, A., Fang, L., Chen, Y., & Chi, M. (2018). New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geoscience and Remote Sensing Magazine*, *6*(3), 10-43.

Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., & Plaza, A. (2017). Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 37-78.

Ghorbani, Z., & Behzadan, A. H. (2021). Monitoring offshore oil pollution using multi-class convolutional neural networks. *Environmental Pollution*, *289*, 117884.

Goetz, A. F. (2009). Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote sensing of environment*, *113*, S5-S16.

Gou, T., Li, Y., Liu, B., & Wang, Z. (2025). Detection of offshore oil slicks utilizing a novel joint feature extraction technique that integrates limited hyperspectral image data. *The Photogrammetric Record*, *40*(189), e12526.

Hossain, M. A., Jia, X., & Pickering, M. (2012). Improved feature selection based on a mutual information measure for hyperspectral image classification. 2012 IEEE international geoscience and remote sensing symposium,

Huang, X., Zhang, B., Perrie, W., Lu, Y., & Wang, C. (2022). A novel deep learning method for marine oil spill detection from satellite synthetic aperture radar imagery. *Marine pollution bulletin*, *179*, 113666.

Huang, Y., Yang, Q.-Y., & Chiang, C.-H. (2023). Applying the modified adaptive window sizes into the image segmentation to identify defects from thermal infrared images covered with intensity inhomogeneity. Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023,

Huby, A. A., Sagban, R., & Alubady, R. (2022). Oil spill detection based on machine learning and deep learning: A review. 2022 5th International Conference on Engineering Technology and its Applications (IICETA),

Jha, S., & Dahiya, P. (2022). Impact analysis of oil pollution on environment, marine, and soil communities. In *Advances in oil-water separation* (pp. 99-113). Elsevier.

Kinga, D., & Adam, J. B. (2015). A method for stochastic optimization. International conference on learning representations (ICLR),

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.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Li, Y., Lyu, X., Frery, A. C., & Ren, P. (2021). Oil spill detection with multiscale conditional adversarial networks with small-data training. *Remote Sensing*, *13*(12), 2378.

Licciardi, G. A. (2019). Hyperspectral compression. In *Data handling in science and technology* (Vol. 32, pp. 55-67). Elsevier.

Liu, B., Li, Y., Liu, C., Xie, F., & Muller, J.-P. (2018). Hyperspectral features of oil-polluted sea ice and the response to the contamination area fraction. *Sensors*, *18*(1), 234.

Liu, B., Li, Y., Zhang, Q., & Han, L. (2016). Assessing sensitivity of hyperspectral sensor to detect oils with sea ice. *Journal of Spectroscopy*, *2016*(1), 6584314.

Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

Massoura, S., & Sommerville, M. (2009). Realities of Spill Response: Waste Remediation Strategies. SPE Asia Pacific Health, Safety, Security, Environment and Social Responsibility Symposium?,

Nieto-Hidalgo, M., Gallego, A.-J., Gil, P., & Pertusa, A. (2018). Two-stage convolutional neural network for ship and spill detection using SLAR images. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(9), 5217-5230.

Park, M.-H., Cho, J.-H., & Kim, Y.-T. (2023). CNN model with multilayer ASPP and two-step cross-stage for semantic segmentation. *Machines*, *11*(2), 126.

Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. (2022). Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, *10*(2), 172-200.

Prasad, G., & Anuprakash, M. (2016). Pollution due to oil spills in marine environment and control measures. *IOSR J. Environ. Sci. Toxicol. Food Technol*, *10*(9), 01-08.

Prasad, K. S. R., Edukondalu, P., & Rao, G. S. (2022). Comparative study and utilization of best deep learning algorithms for the image processing. *ResearchGate*, *18*, 227.

Prechelt, L. (2002). Early stopping-but when? In *Neural Networks: Tricks of the trade* (pp. 55-69). Springer.

Qian, S. E. (2022). Overview of hyperspectral imaging remote sensing from satellites. *Advances in Hyperspectral Image Processing Techniques*, 41-66.

Rahman, M. A., & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. International symposium on visual computing,

Rajendran, S., Vethamony, P., Sadooni, F. N., Al-Kuwari, H. A.-S., Al-Khayat, J. A., Seegobin, V. O., Govil, H., & Nasir, S. (2021). Detection of Wakashio oil spill off Mauritius using Sentinel-1 and 2 data: Capability of sensors, image transformation methods and mapping. *Environmental Pollution*, *274*, 116618.

Raju, T., Narayan, A., Abinayaa, R., & Joseph, C. (2024). Detection Of Oil Spill in Satellite-Based Synthetic Aperture Radar Images by Neural Network. 2024 4th Asian Conference on Innovation in Technology (ASIANCON),

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18,

Satyanarayana, A. R., & Dhali, M. A. (2023). Oil Spill Segmentation using Deep Encoder-Decoder models. *arXiv preprint arXiv:2305.01386*.

Shabna, A., & Ganesan, R. (2014). HSEG and PCA for Hyper-spectral Image Classification. 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT),

Shanmukh, M. P., Priya, S. B., & Madeswaran, T. (2024). Improving oil spill detection in marine environments through deep learning approaches. 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT),

Sivadas, S. K., Muthukumar, C., Bharathi, M., Ramu, K., Srivastava, P. K., & Murthy, M. R. (2021). Connecting India's coastal monitoring program with UN Sustainable Development Goal 14. *Ocean & Coastal Management*, *215*, 105949.

Uddin, M. P., Mamun, M. A., & Hossain, M. A. (2019). Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. *International Journal of Remote Sensing*, *40*(18), 7190-7220.

Ustuner, M. (2024). Randomized Principal Component Analysis for Hyperspectral Image Classification. 2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS),

Vijayakumar, S. (2023). Computational Techniques of Oil Spill Detection in Synthetic Aperture Radar Data: Review Cases. *Recent Oil Spill Challenges That Require More Attention*.

Wang, X., Liu, J., Zhang, S., Deng, Q., Wang, Z., Li, Y., & Fan, J. (2021). Detection of oil spill using SAR imagery based on AlexNet model. *Computational Intelligence and Neuroscience*, *2021*(1), 4812979.

Xie, M., Li, Y., Zhang, Z., Fu, Q., & Jiang, H. (2025). Remote sensing of the oil spills caused by ships: A review. *Marine pollution bulletin*, *214*, 117754.

Yekeen, S. T., Balogun, A. L., & Yusof, K. B. W. (2020). A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, *167*, 190-200.

Zhao, D., Tan, B., Zhang, H., & Deng, R. (2022). Monitoring marine oil spills in hyperspectral and multispectral remote sensing data by the spectral gene extraction (SGE) method. *Sustainability*, *14*(20), 13696.

Table 1. Overview of HOSD Image Dataset: Spatial Size, Flight Time, and Resolution

Image name	Spatial size	Fight time	Resolution(m)
GM1	1200×663	5/17/2010	7.6
GM2	1881×693	5/17/2010	7.6
GM3	1430×691	5/17/2010	7.6
GM4	1700×691	5/17/2010	7.6
GM5	2042×673	5/17/2010	7.6
GM6	2128×689	5/18/2010	8.1
GM7	2302×479	7/09/2010	3.3
GM8	1668×550	7/09/2010	3.3
GM9	1643×447	7/09/2010	3.2
GM10	1110×675	5/17/2010	7.6
GM11	1206×675	5/17/2010	7.6
GM12	869×649	5/06/2010	7.6
GM13	1135×527	7/09/2010	3.2
GM14	1790×527	7/09/2010	3.2
GM15	1777×510	7/09/2010	3.3
GM16	1159×388	7/09/2010	3.2
GM17	1136×660	5/17/2010	7.6
GM18	1047×550	7/09/2010	3.3

Table 2. Comparison of model size and parameter count for U-Net and DeepLabv3

model	Number of parameters	Total size (MB)

U-Net	12567522	47.94
DeepLabv3	26861986	102.47

Table 3. Comparison of segmentation models in terms of evaluation metrics (%)

Model name	**Accuracy**	**Sensitivity**	**Precision**	**IoU**	**F1-score**
U-Net	96.67	63.34	76.56	53	69.32
DeepLabv3	97.04	68.54	78.64	57.86	73.30

Fig. 1. Location of the Deepwater Horizon oil spill in the Gulf of Mexico

Fig. 1. Location of the Deepwater Horizon oil spill in the Gulf of Mexico

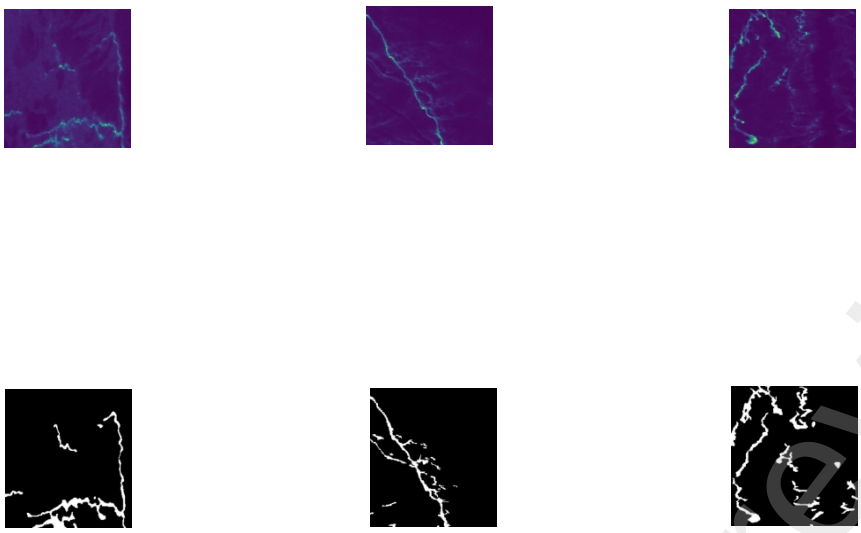


Fig. 2. Three representative examples of the 256 × 256 pixel original images and their corresponding ground truth masks from the HOSD dataset.

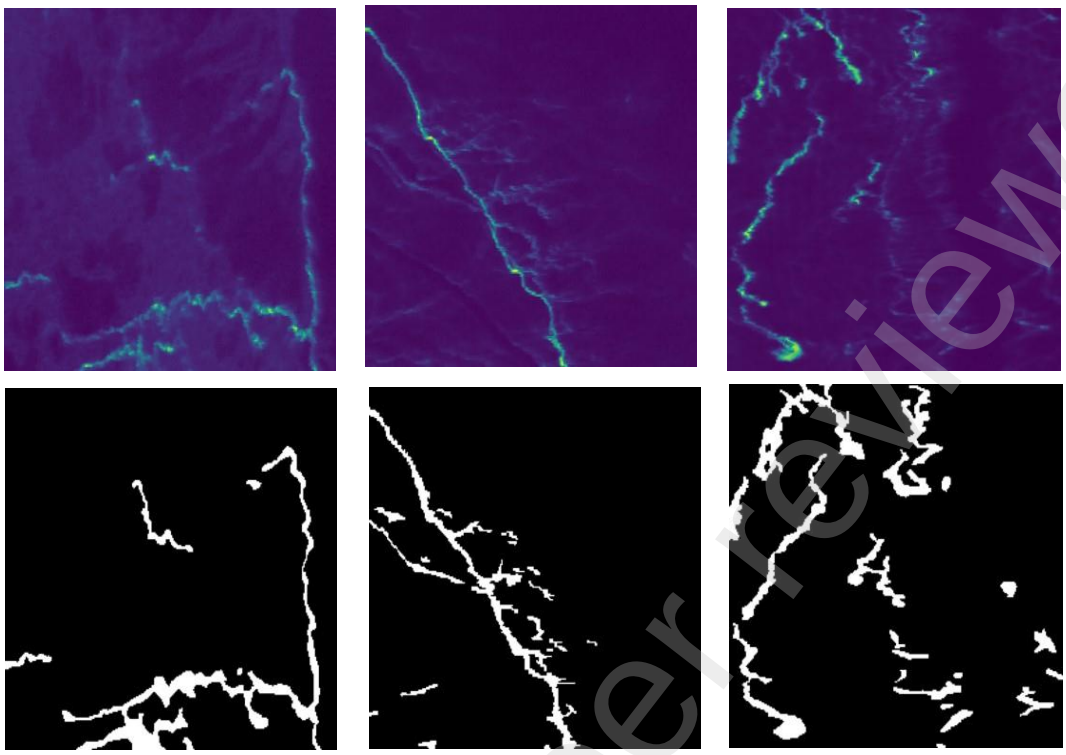


Fig. 2. Three representative examples of the 256 × 256 pixel original images and their corresponding ground truth masks from the HOSD dataset.

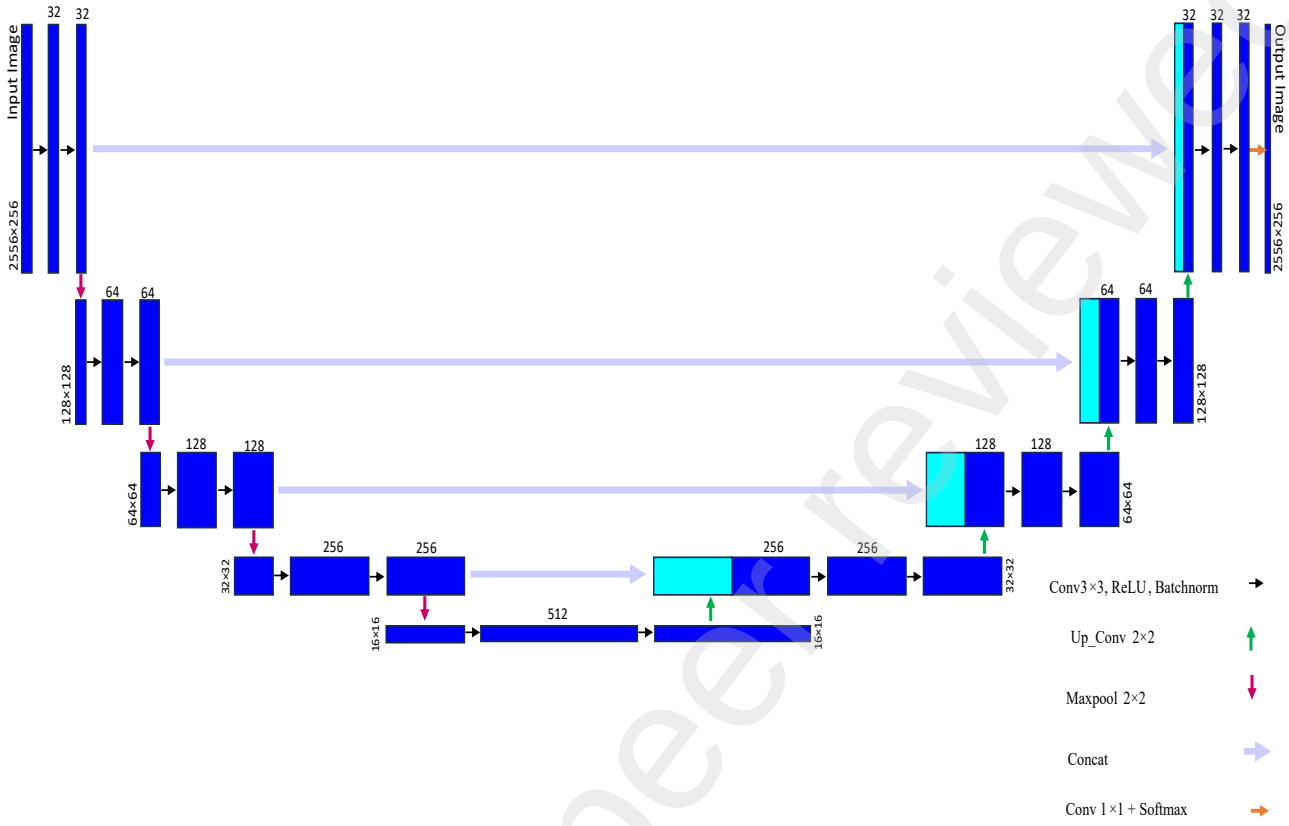


Fig. 3. Architecture of the U-Net used in this work

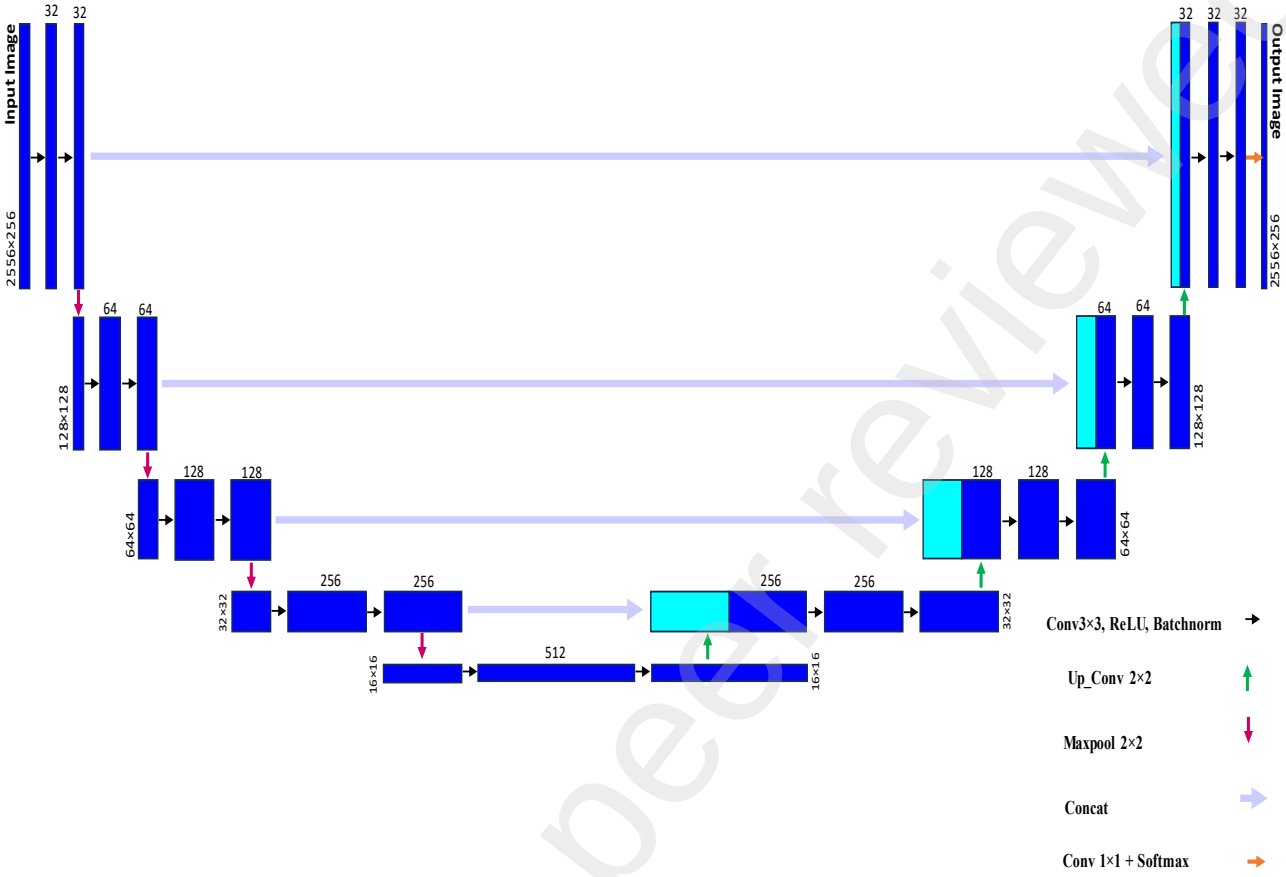


Fig. 3. Architecture of the U-Net used in this work

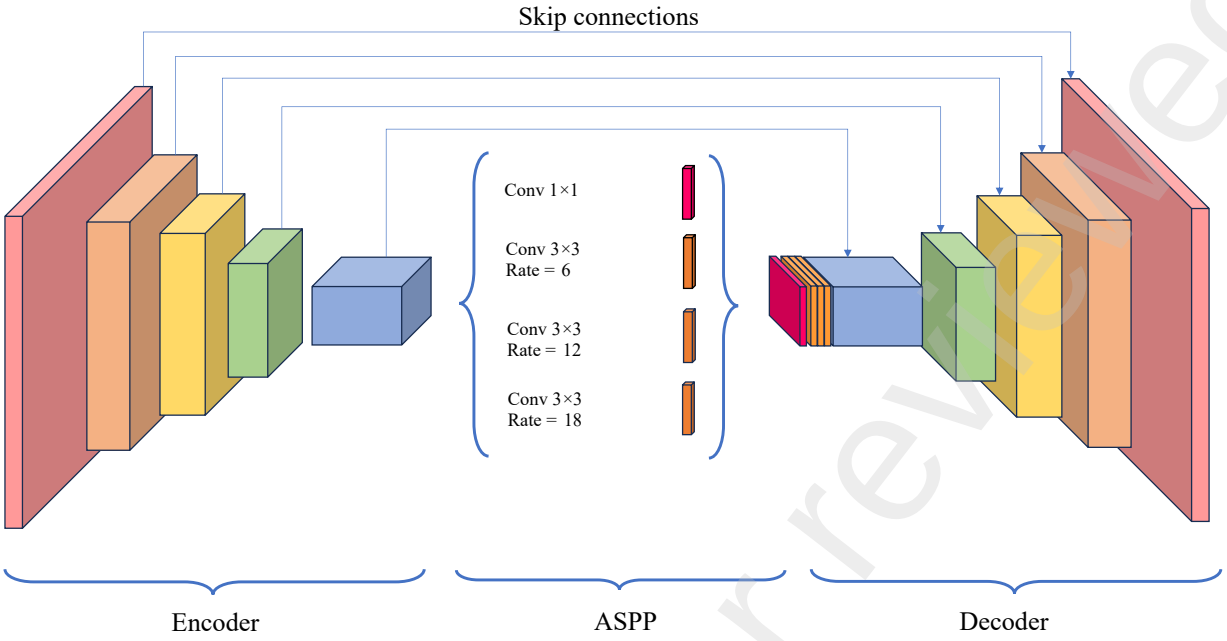


Fig. 4. Architecture of the deeplabv3 used in this work

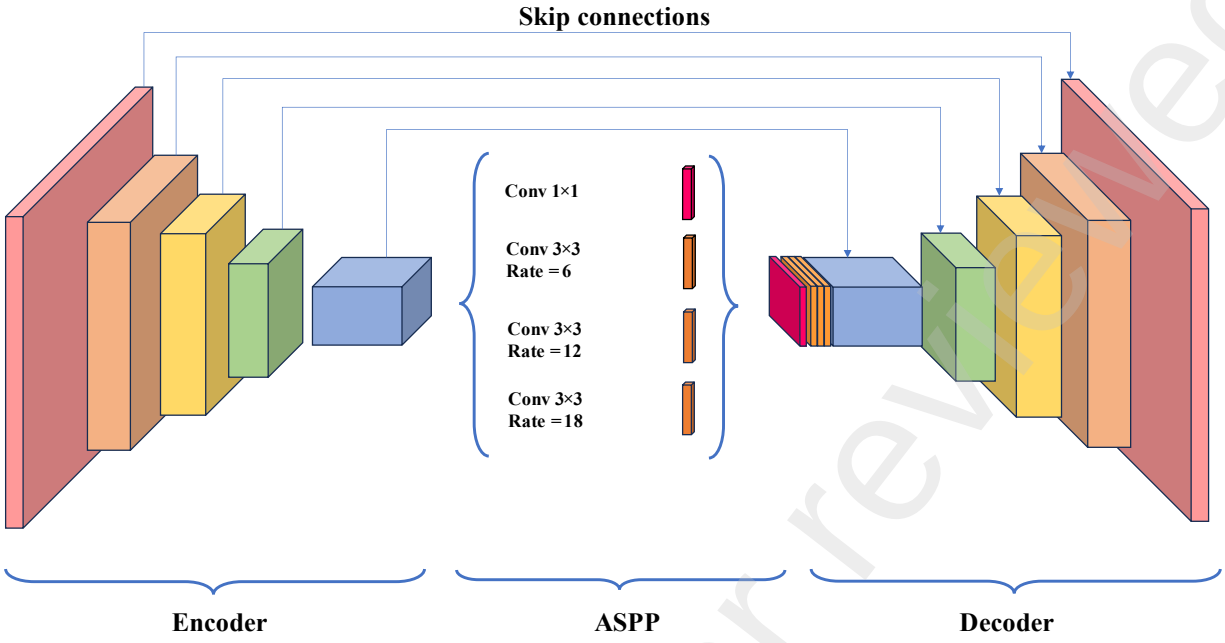


Fig. 4. Architecture of the deeplabv3 used in this work

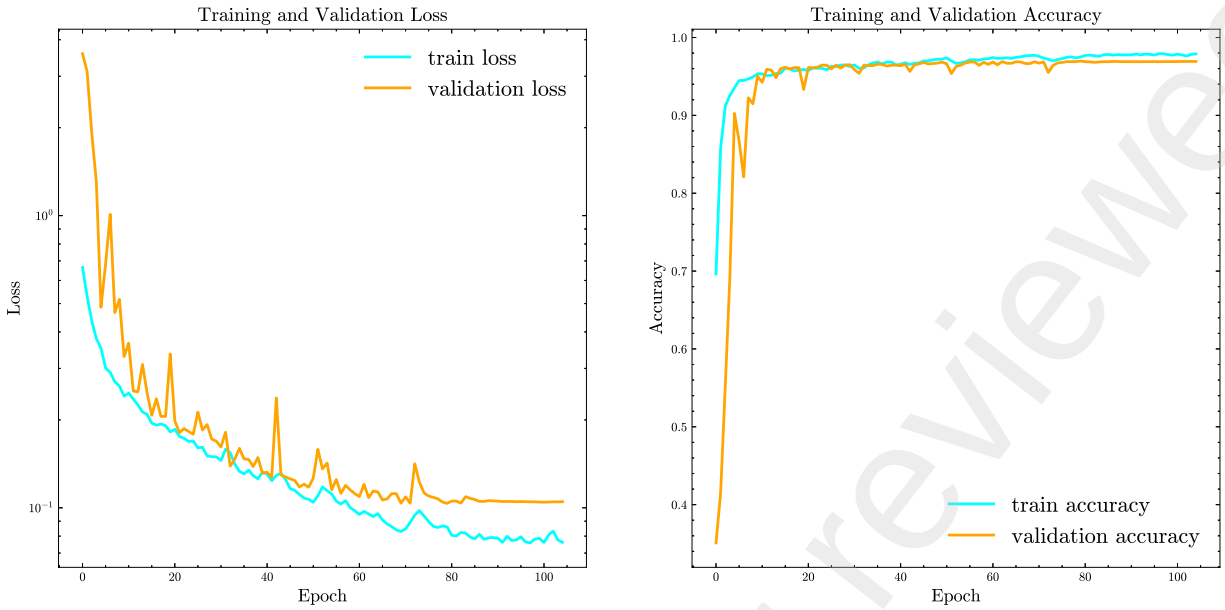


Fig. 5. Training and validation loss and accuracy for the DeepLabV3 Oil Spill Detection Model

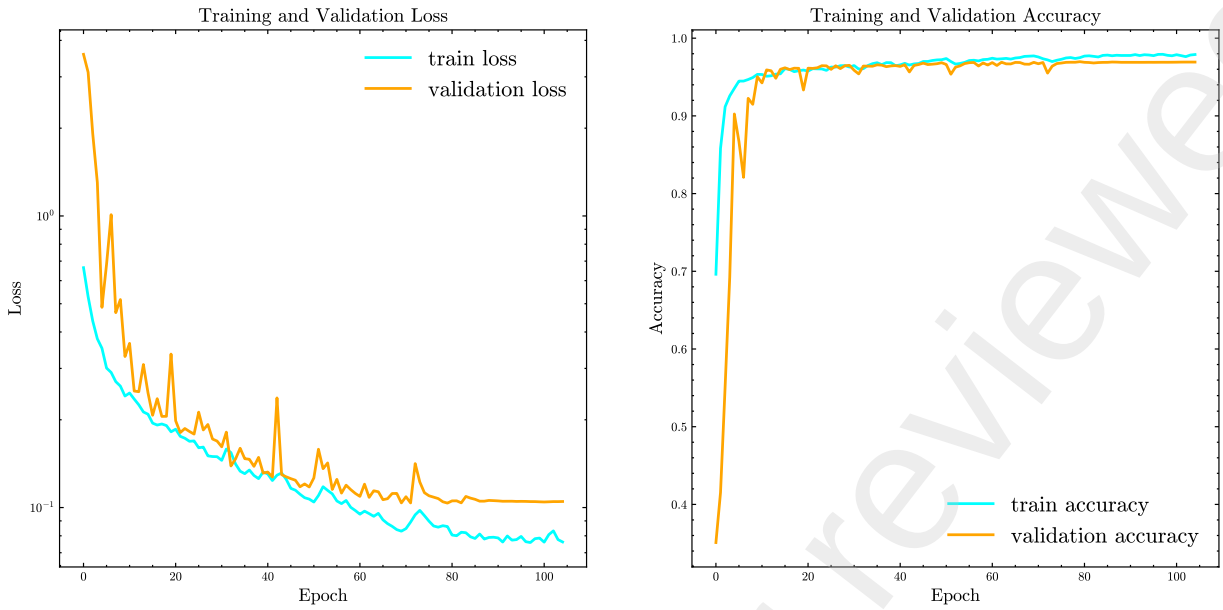
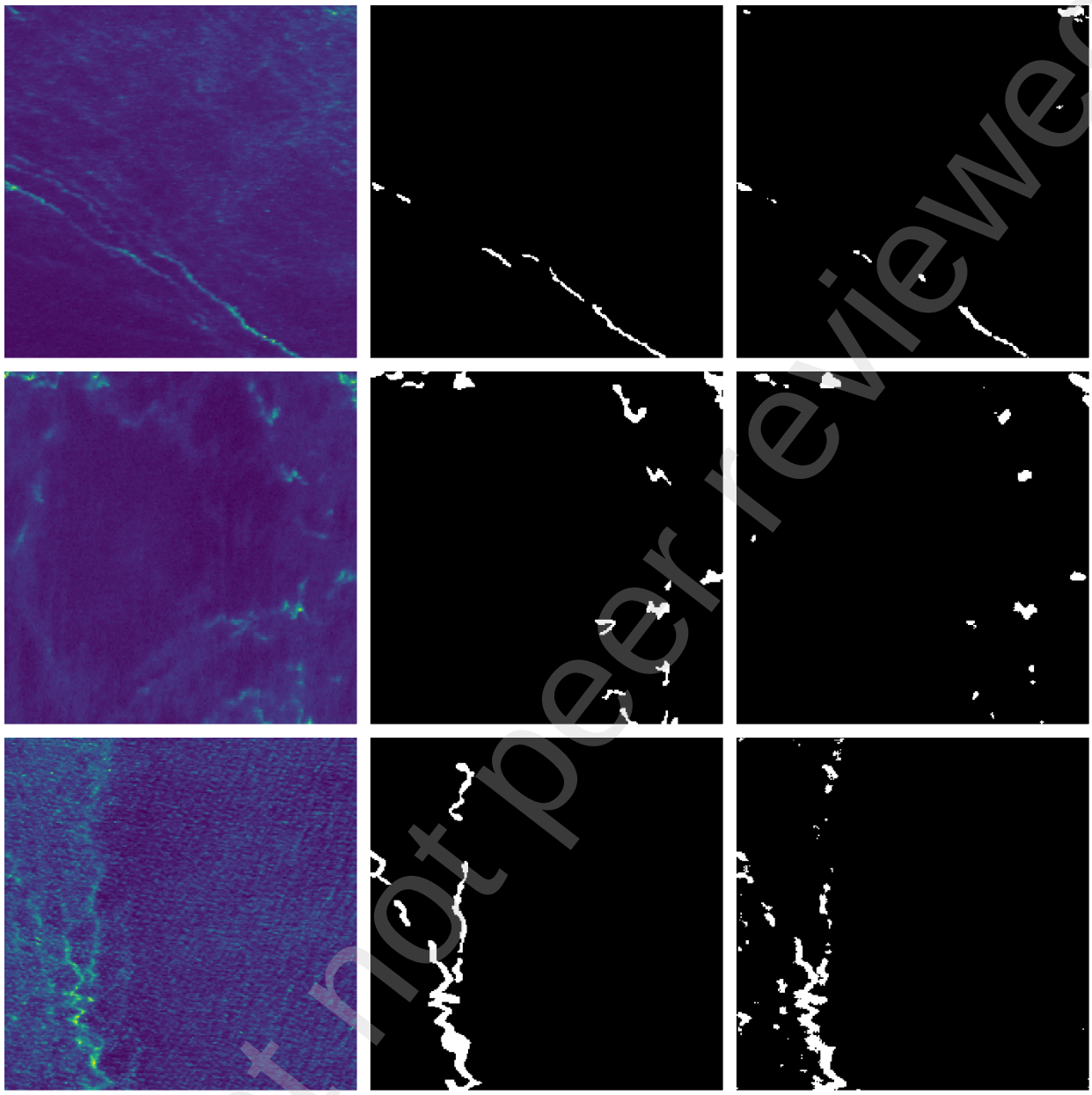


Fig. 5. Training and validation loss and accuracy for the DeepLabV3 Oil Spill Detection Model



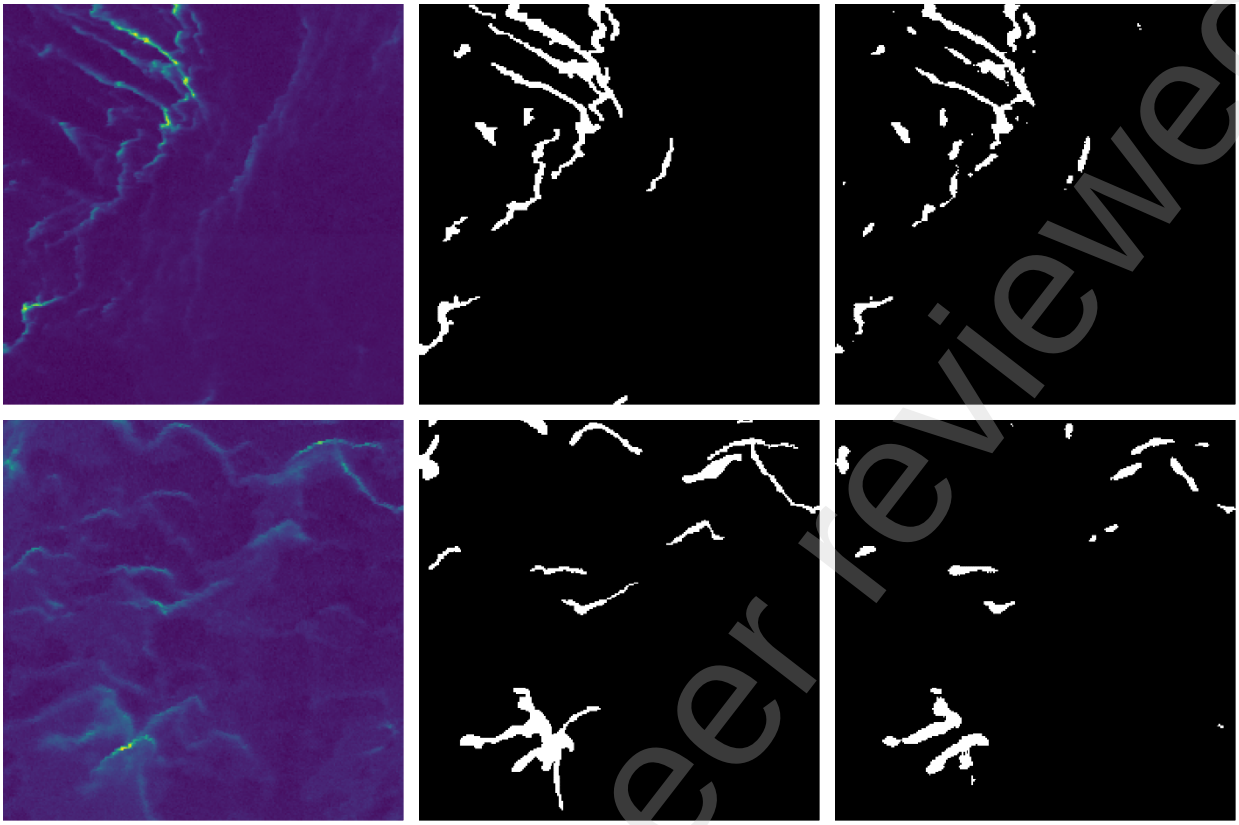
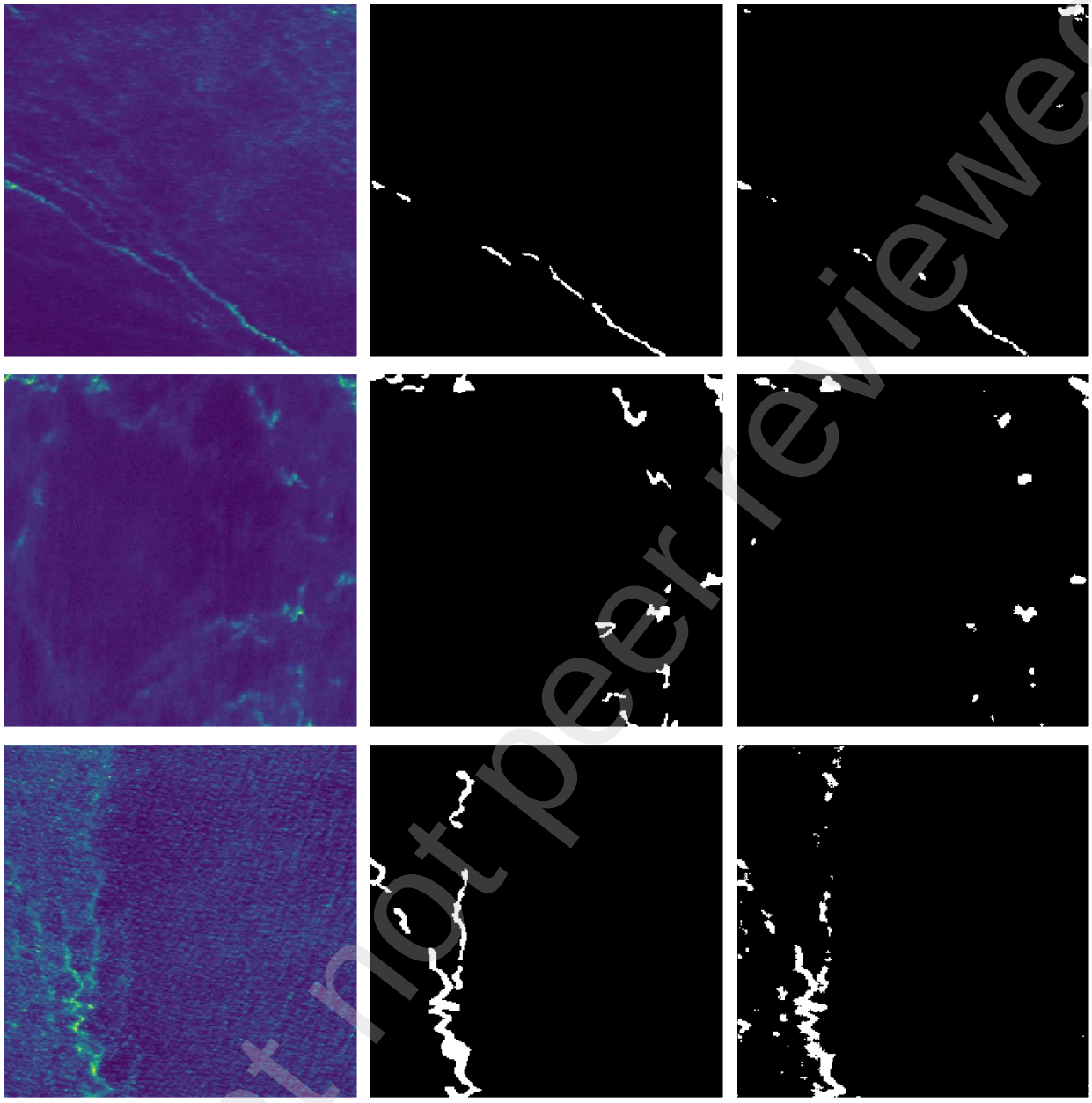


Fig. 6. shows five examples of detected oil spills with Deeplabv3

Original Image	Ground Truth	Prediction



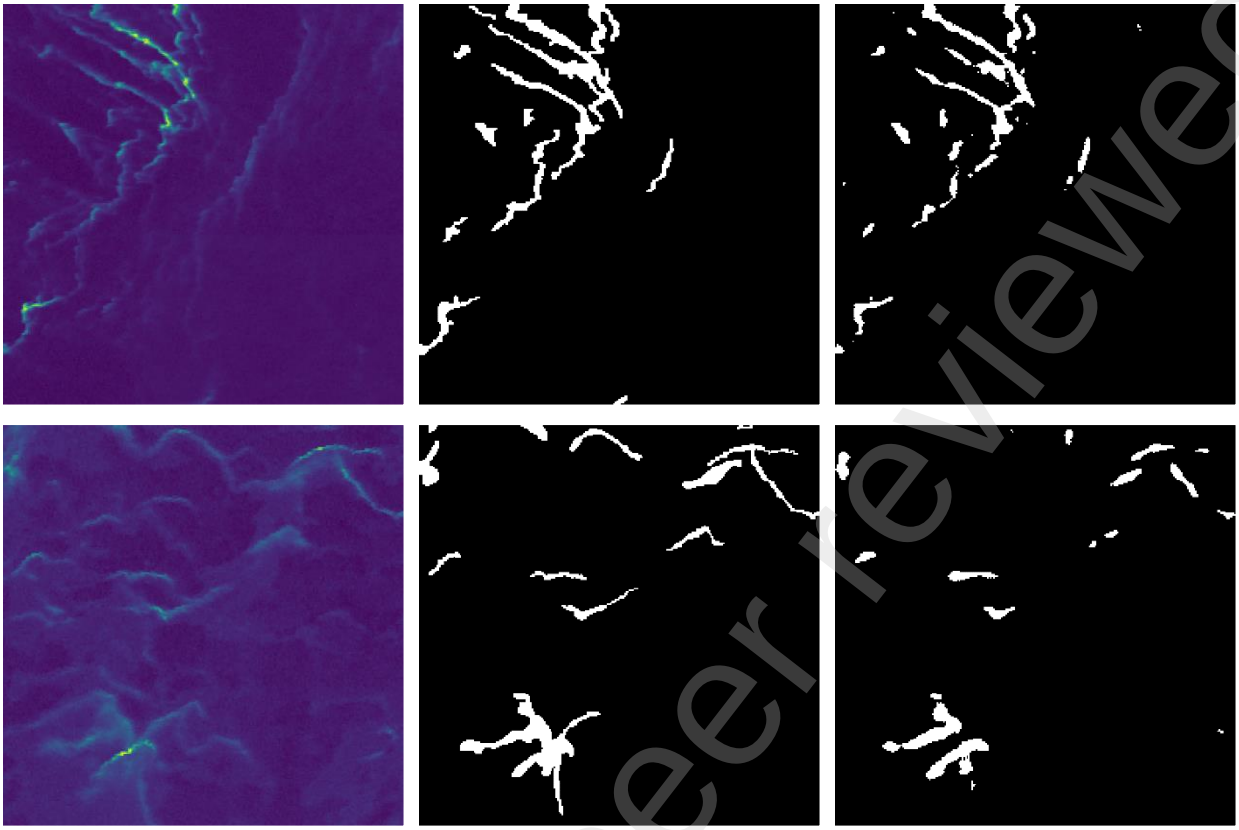


Fig. 6. shows five examples of detected oil spills with Deeplabv3