

Article

A Novel Multi-Scale Feature Map Fusion for Oil Spill Detection of SAR Remote Sensing

Chunshan Li ¹, Yushuai Yang ¹, Xiaofei Yang ^{2,*}, Dianhui Chu ¹ and Weijia Cao ³ 

¹ School of Computer Science and Technology, Harbin Institute of Technology, Weihai 264209, China; lics@hit.edu.cn (C.L.); yangyushuai1016@gmail.com (Y.Y.); chudh@hit.edu.cn (D.C.)

² School of Electronic and Communication Engineering, Guangzhou University, Guangzhou 510006, China

³ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; caowj@aircas.ac.cn

* Correspondence: xiaofeiyang@gzhu.edu.cn

Abstract: The efficient and timely identification of oil spill areas is crucial for ocean environmental protection. Synthetic aperture radar (SAR) is widely used in oil spill detection due to its all-weather monitoring capability. Meanwhile, existing deep learning-based oil spill detection methods mainly rely on the classical U-Net framework and have achieved impressive results. However, SAR images exhibit high noise, blurry boundaries, and irregular shapes of target areas, as well as speckles and shadows, which lead to the loss of performance in existing algorithms. In this paper, we propose a novel network architecture to achieve more precise segmentation of oil spill areas by reintroducing rich semantic contextual information before obtaining the final segmentation mask. Specifically, the proposed architecture can re-fuse feature maps from different levels at the decoder end. We design a multi-convolutional layer (MCL) module to extract basic feature information from SAR images, and a feature extraction module (FEM) module further extracts and fuses feature maps generated by the U-Net decoder at different levels. Through these operations, the network can learn rich global and local contextual information, enable sufficient interaction of feature information at different stages, enhance the model’s contextual awareness, and improve its ability to recognize complex textures and blurry boundaries, thereby enhancing the segmentation accuracy of SAR images. Compared to many U-Net based segmentation networks, our method shows promising results and achieves state-of-the-art performance on multiple evaluation metrics.



Citation: Li, C.; Yang, Y.; Yang, X.; Chu, D.; Cao, W. A Novel Multi-Scale Feature Map Fusion for Oil Spill Detection of SAR Remote Sensing. *Remote Sens.* **2024**, *16*, 1684. <https://doi.org/10.3390/rs16101684>

Academic Editor: Piotr Samczynski

Received: 31 March 2024

Revised: 6 May 2024

Accepted: 7 May 2024

Published: 9 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The ocean comprises a vital component of our global ecosystem, exerting a profound influence on global climate and environmental shifts as evidenced by numerous studies [1–3]. However, marine pollution has escalated into a pressing global concern [4–6], particularly in the context of oil spills occurring during maritime transportation [7,8]. Oil spills have the potential to blanket the water’s surface, impeding sunlight penetration and consequently impairing the photosynthetic processes of aquatic flora. This disruption can lead to the demise of aquatic vegetation, which subsequently undermines the stability of the marine food chain. The consequences are not limited to marine life; they also encompass critical ecosystem functions [9,10]. Furthermore, the presence of toxic chemicals in spilled oil can cause direct fatality or reproductive impairments in aquatic fauna, such as fish, when accidentally ingested [11,12]. The resulting substantial decline in their populations has significant ramifications for the fishing industry. Simultaneously, hydrocarbons released during oil spills accumulate in the atmosphere, contributing to the formation of greenhouse gases that trap heat and exacerbate the greenhouse effect. This, in turn, has far-reaching impacts on human society. Hence, the repercussions of oil spills are

widespread and profound, affecting the environment, economy, ecology, and public health on a significant scale [13,14].

Oil spills often stem from incidents involving ships or tankers, ruptured pipelines, illicit discharges, and the improper disposal of bilge oil residue. Furthermore, accidents on offshore drilling platforms or within petroleum pipelines can also result in significant oil leaks [15–17]. Statistics reveal that approximately 380 million gallons of oil are spilled annually due to natural disasters and anthropogenic activities. Notably, human factors contribute to roughly 45% of global oil-related environmental pollution from spills, with transportation-related oil leakage accounting for 5% [18,19]. Oil possesses remarkable persistence, exhibiting a sluggish degradation rate that necessitates a considerable amount of time for natural decomposition. Consequently, the environmental implications of an oil spill can persist for an extended period, posing significant challenges for ecosystem recovery and restoration, especially in intricate environments like the oceans.

Marine oil spill detection is considered a critically important task in ocean observation. Effective oil spill monitoring not only aids in the early detection and tracking of potential pollution sources but also guides emergency responses and measures to alleviate the adverse impact of oil spill incidents on the environment, economy, and society. Traditional detection methods require on-site human identification, but such direct contact with oil poses safety risks [20]. Subsequently, marine detection systems have utilized aircraft and coast guard units for inspections. While these methods can accomplish the task of oil spill detection, the issue of high costs has emerged as a concern [21]. With the continuous advancement of remote sensing technology, people are gradually realizing that using satellites for marine oil spill detection is a readily available, cost-effective, and low-risk method. This approach is capable of accurately capturing signs of oil spills, aiding in issuing early warnings and responding rapidly to potential pollution incidents. By obtaining real-time data from vast oceanic areas, we can swiftly identify oil spill situations and take precise measures to mitigate potential environmental and economic risks.

Recently, synthetic aperture radar (SAR) satellite remote sensing has become the primary method for locating maritime oil spills. This technology utilizes synthetic aperture radar carried by satellites to efficiently detect oil films on the ocean surface, without being limited by weather or lighting conditions [22,23]. SAR satellite remote sensing has numerous advantages compared to technologies such as hyperspectral remote sensing, visible light remote sensing, and thermal infrared remote sensing [24–26]. Firstly, SAR satellites utilize radar beams to transmit and receive electromagnetic waves, unaffected by weather and lighting conditions. They can provide reliable remote sensing data in all weather conditions, including daytime, nighttime, clear weather, and cloudy weather. On the contrary, technologies such as hyperspectral remote sensing, visible light remote sensing, and thermal infrared remote sensing are limited by weather and lighting conditions. Factors such as cloud cover and atmospheric humidity can cause a significant decline in the quality of monitoring images. Secondly, SAR satellite remote sensing technology has higher spatial resolution, which is advantageous for locating oil spill areas, identifying boundaries, and assessing the extent of oil spills. Meanwhile, visible light remote sensing and thermal infrared remote sensing technologies generally have lower spatial resolution, which imposes limitations on the accuracy of subsequent oil spill area detection.

In addition, certain other phenomena on the sea surface can also cause dark areas in SAR images [27–29], including areas with low wind speed, internal waves in the water, organic films or biological substances on the water surface, the formation of thin oil slicks, rainfall, and ocean currents. Simultaneously, SAR images exhibit characteristics such as high noise levels, unclear boundary definition, and uneven intensity distribution. These factors could potentially impact the accurate positioning and identification of oil spill features within the images. A set of SAR example images is depicted in Figure 1. The oil spill areas displayed in SAR images exhibit irregular shapes, often accompanied by spots, shadows, and various textures. These visual characteristics could be indicative of oil spills,

but they could also be influenced by factors such as marine environment, waves, and wind, which motivates us to develop more effective algorithms for oil spill detection.

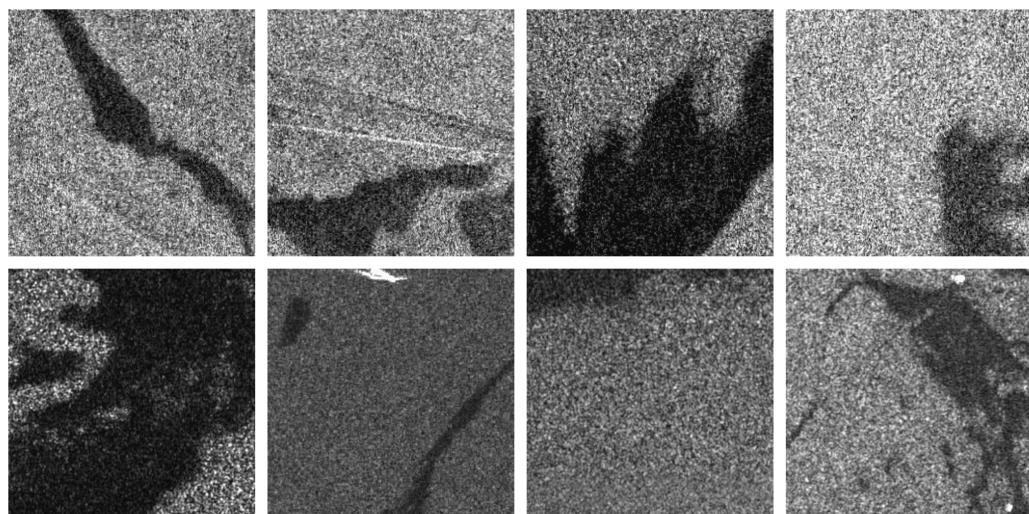


Figure 1. SAR example imagery. They come from the PALSAR dataset and the SENTINEL dataset, which are the two main remote sensing data resources provided by SAR remote sensing technology. The darker areas indicate potential oil spill zones, while the rest of the areas are non-spill zones.

Fortunately, with the continuous improvement of hardware capabilities, deep learning technology has made significant advancements, showcasing its immense potential in the field of image recognition. Generally speaking, the oil spill detection task can be classified as an image semantic segmentation problem, wherein each pixel in the image is divided into two categories. One category represents the oil spill area, while the other represents the non-oil spill area. In this task, the objective is to assign accurate semantic labels to every pixel in the image, in order to precisely distinguish between the oil spill and non-oil spill areas. Deep learning models possess exceptional feature extraction and learning capabilities. Leveraging similar datasets during the training phase, these models can effectively synthesize information inherent in the data and facilitate comprehensive end-to-end analysis. This methodology obviates the need for the laborious manual feature engineering steps associated with traditional approaches, enabling the model to autonomously learn the most discriminative and expressive features from the data.

As a highly efficient visual feature extraction approach, convolutional neural networks (CNNs) are widely applied in the field of image semantic segmentation. By adjusting network architectures, numerous variant models have emerged. These variant models consistently explore innovations in various aspects such as hierarchical structures, loss functions, and skip connections, with the aim of further enhancing the performance of segmentation tasks. The FCN [30] employs fully convolutional layers for semantic segmentation, eliminating fully connected layers. This not only reduces the parameter count but also allows the model to take in input images of arbitrary sizes and produce segmentation masks of the same dimensions. This enhances the model's flexibility and generalization capability. SegNet [31] adopts an encoder-decoder architecture, leveraging multiple layers of convolution and pooling operations to progressively reduce the size of feature maps and extract higher-level features. During the encoding phase, features are extracted by the encoder, while the decoding phase utilizes the decoder to restore features to the input size. This approach aids in recovering image details, enabling pixel-level segmentation predictions. U-Net [32] was originally applied in the field of medical image segmentation, and it offers a richer acquisition of multi-scale information compared to FCN. The uniqueness of U-Net lies in its U-shaped architecture, which combines encoders and decoders, enabling the simultaneous capture of both local and global features within images. This allows U-Net to more accurately recover details and perform finer image segmentation.

Derived networks from U-Net include Unet++ [33], R2Unet [34], TransUnet [35], AttU-Net [36] and Swin-Unet [37], among others. These variants build upon the foundation of U-Net and introduce varying degrees of improvements and extensions. For instance, Unet++ employs a nested architecture enhancement, Swin-Unet integrates the Swin Transformer to enhance feature extraction, R2Unet introduces a recurrent mechanism, TransUnet combines Transformer modules, and AttU-Net introduces attention mechanisms. These additions aim to further optimize feature learning and image segmentation performance within the networks, enabling these variant architectures to better address diverse types of image segmentation tasks.

Although these U-Net networks can aggregate features into the deepest feature map produced by convolution operations, they may ignore rich contextual information contained in the different-scale feature maps, which may lead to loss of performance. To solve the above issues, in this paper, we design a novel framework aimed at addressing oil spill detection tasks. Our proposed model assumes that within the decoding phase of U-Net, the intermediate layers encompass a substantial amount of information that is crucial for generating the final segmentation mask. The presence of this information enables the network to capture the subtle variations in different feature regions of the image more accurately during segmentation tasks, thereby enhancing the precision and reliability of the segmentation outcomes. Compared to the previous methods [30–37], our model achieves promising high scores on metrics: Dice score, HD95, precision, and accuracy. This indicates that our model demonstrates excellent performance across multiple key evaluation criteria.

The main contributions of this article are as follows:

1. We propose a novel network architecture which extends U-Net framework. By integrating different scale feature information generated at various stages of the decoder, the model effectively utilizes multi-stage contextual information, thereby significantly enhancing the accuracy of oil spill area identification in oil spill segmentation tasks.
2. A multi-convolutional layer (MCL) module is introduced to extract crucial feature information from the input SAR image. By employing the MCL module, we utilize transpose convolution operations to increase the image dimensions. In contrast to regular convolution, the MCL enlarges the size of feature maps during the decoding phase and aids in recovering lost details and spatial resolution. Simultaneously, it effectively remaps abstract features from the encoder to the decoder, ensuring comprehensive and accurate feature representation.
3. A feature extraction module (FEM) is carefully designed to integrate feature maps of varying scales. This module incorporates a channel attention mechanism that can adaptively adjust the weights of feature maps along the channel dimension. This process enhances the ability of the proposed model to understand information from different channels.
4. We conduct extensive experiments on two distinct SAR datasets to verify the performance of the proposed model. And the results show that, compared to existing semantic segmentation networks, our proposed model achieves the current state-of-the-art performance.

The remaining sections of this paper are organized as follows: Section 2 introduces the relevant works about oil spill detection. Section 3 describes the network structure we propose. In Section 4, we conduct comparative experiments by training different models on the same datasets, and perform ablation studies on our model architecture. Subsequently, we discuss and analyze the results of these experiments. Finally, in Section 6, we present the conclusions of our study and outline the prospects for future research directions.

2. Related Work

In the domain of SAR satellite remote sensing for oil spill detection, there are two primary research directions. The first one relies on traditional image processing techniques and feature extraction methods including thresholding, edge detection, texture analysis, and region growing. These methods extract features such as edges and textures from the

image to achieve target segmentation and recognition. The second direction leverages deep neural network models, which can learn high-level latent representations and semantic information of the images aiding segmentation and recognition. These deep neural network models can automatically learn complex features and semantic information from the images, thereby improving the accuracy and robustness of segmentation and recognition.

2.1. Traditional Image Processing Techniques

In earlier studies, researchers widely employed traditional image processing methods, including thresholding, edge detection, texture analysis, and region growing. Literature [38] proposed a dual-threshold oil spill image segmentation algorithm, which utilizes high and low thresholds to extract grayscale information at different levels and performs segmentation of oil spill regions based on a feature probability function. Ref. [39] presents a novel oil spill feature selection and classification technique based on a forest of decision trees. It adopts a framework of a multi-objective problem, aiming to minimize the used input features while maximizing the overall testing classification accuracy. A self-adaptive mechanism based on the Otsu method is proposed in [40], combining region growing with edge detection and threshold segmentation (RGEDOM) for oil spill extraction. In [41], the classification algorithm combines classification tree analysis and fuzzy logic to process the data in two stages. Feature parameters are extracted from each segmented dark spot for oil spill and 'look-alike' classification, and ranked according to their importance. In [42], an improved edge detection algorithm based on the Canny algorithm and thresholding algorithm is proposed to address the issue of high standard deviation statistical distribution caused by speckle noise in SAR images. In [43], a semi-automatic oil spill detection method is proposed that does not require manual threshold setting, enabling the extraction of oil spills in a semi-automated manner. The method utilizes texture analysis, machine learning, and adaptive thresholding on X-band marine radar images. By combining the region growing method and multi-scale analysis algorithm, Ref. [44] proposes a technique for locating dark areas in the ocean that may potentially be oil slicks. The method utilizes multi-scale analysis with undecimated wavelets to smooth the speckle noise in SAR images and enhance edges. In the field of traditional machine learning algorithms, the support vector machine (SVM) method and artificial neural network (ANN) technique are also being used to detect and recognize oil spill areas. Ref. [45] proposes a generic and systematic approach based on machine learning-based feature selection methods to select concise and relevant feature sets for improving oil spill detection systems. Ref. [46] combines the object-based classification method with Support Vector Machines (SVMs) for oil spill identification. This method classifies dark patches in SAR images to distinguish oil spill areas from other similar phenomena. Ref. [47] utilizes artificial neural networks (ANNs) for image segmentation and classification to distinguish oil spill areas from similar regions. The core idea is to employ two separate ANNs in a sequential manner. The first ANN is employed for SAR image segmentation, identifying candidate pixels representing oil spill features. Subsequently, statistical feature parameters are extracted and used to drive the second ANN, which classifies the targets as either oil spills or similar phenomena. However, the above methods often require the manual design and selection of feature extractors, which are time-consuming and dependent on expert knowledge, and often optimize for specific scenarios or datasets, with relatively weak generalization ability. Therefore, these methods have encountered performance bottlenecks in practical applications. Ref. [48] conducts in-depth research on the characteristics of oil spills and divides them into three categories: geometric features, physical behavior features, and contextual features. Geometric features, such as area, perimeter, and complexity, describe the shape and size of the oil pollution. Physical behavior features, such as average or maximum echo value, standard deviation of dark areas, or size of surrounding areas, capture the physical properties and behavior of the oil pollution. Contextual features provide information about the context of oil pollution in the image, such as the number of other dark areas or the presence of vessels. Finally, a genetic algorithm is used to identify the feature combinations that are

most relevant to oil spill detection, helping to improve the accuracy of oil spill detection. Ref. [49] employs evolutionary algorithms and Bayesian network methods to investigate the features of SAR images. The author extracts various features from the SAR images, encompassing geometric, texture, physical, and contextual aspects. In order to determine the optimal subset of features, the author utilizes eight distinct evolutionary algorithms and assesses the classification error rates of the generated feature sets using Bayesian networks. The primary objective is to enhance the precision in discriminating between oil spills and similar substances.

2.2. Advancements in Deep Learning Technologies

In recent years, deep learning techniques have been receiving widespread attention and research from numerous scholars due to their end-to-end training approach and powerful learning capabilities. Ref. [50] proposes a feature fusion network (FMNet) for the segmentation of oil spill areas in SAR images. It utilizes a threshold segmentation method to obtain the global features of the SAR image. High-dimensional features are then extracted from the threshold segmentation results using convolutional operations, enabling the model to make more accurate decisions. Ref. [51] addresses the issue of the underutilization of phase information and other polarimetric information in SAR images by proposing an intelligent oil spill detection architecture based on a deep convolutional neural network (DCNN). Ref. [52] proposes a Deeplabv3+ semantic segmentation algorithm that utilizes multiple loss function constraints and a multi-level cascading residual structure. Ref. [53] addresses the problem of training an oil spill detection model with limited data by proposing a multi-scale conditional adversarial network (MCAN) comprised of adversarial networks at multiple scales. The multi-scale architecture comprehensively captures both global and local oil spill characteristics, while adversarial training enhances the model's representational power through generated data. This paper is based on deep learning neural network technology and proposes a simple and efficient multi-scale fusion strategy for the feature maps generated at different stages of the decoder in the U-Net architecture. The strategy involves the re-extraction and fusion of feature maps from different stages, enabling interaction between contextual information at each stage. This approach provides a more comprehensive and rich feature representation, leading to better segmentation results for oil spill regions. In summary, deep learning has gradually replaced traditional methods in the field of image segmentation due to its powerful feature learning ability, end-to-end learning approach, powerful generalization ability, improved computing resources, and continuous progress of algorithms.

3. Methods

In this section, we provide a detailed exposition of the information regarding the SAR dataset we have employed, along with unveiling the core principles and model architecture of our proposed method.

3.1. Overall Framework

As U-Net [32] demonstrates remarkable performance in the field of image segmentation, an increasing number of variant networks based on U-Net are emerging [33–37]. With the continuous in-depth research into the U-Net architecture, we are increasingly confident that the decoder-side intermediate feature maps of the U-Net network architecture may contain abundant contextual information that is highly beneficial for the final segmentation mask. These feature maps possess multiple scales, with each scale presenting unique characteristics of the image. They have the capability to capture both the global structure and subtle details of the image. In the segmentation process, they provide strong guidance for pixel classification and region segmentation, contributing to a more accurate understanding of the image's structure and details. Therefore, we believe that harnessing the abundant contextual information contained in the intermediate feature maps of the U-Net decoder is crucial for enhancing the accuracy and robustness of image segmentation outcomes.

Building upon the aforementioned considerations, our objective is to delve into the exploration of more effective ways to capture image features, aiming to achieve a more precise segmentation of SAR images. Therefore, we have constructed a model with the aim of validating the effectiveness of our ideas and methods. As depicted in Figure 2, this represents the model we have designed following the U-Net architecture. Firstly, we preprocess the SAR image with dimensions [3, 256, 256], transforming it into a single-channel grayscale image with dimensions [1, 256, 256], and then input it into the network. Secondly, in the encoding phase, the U-Net encoder continually extracts features from the image through multiple layers of convolutional layers while performing downsampling operations on the image multiple times. This process helps to gradually abstract and compress the image information, thereby obtaining higher-level feature representations in preparation for the subsequent decoding phase. With each layer's processing, the semantic information and abstraction level of the image progressively enhance, laying the foundation for more accurate segmentation results. During the downsampling process, we employ convolution operations with a stride of 2, which effectively reduces the image size. Utilizing convolutions with a stride of 2 enhances the learning of image feature representation compared to methods like max pooling. Convolution operations, while downsampling, retain more image details and contextual information, enabling the capture of valuable features within the image. This approach diminishes information loss, allowing the model to more effectively extract task-relevant information from the input image. Thirdly, the feature maps extracted by the encoder undergo processing by the U-Net decoder. During this process, we employ shortcut connections to complement and merge feature maps from different levels of the encoder with corresponding levels in the decoder. This practice facilitates the mutual exchange of high-level semantic information and low-level detail information, thus comprehensively capturing image features and contextual information. Within this process, the upsampling step employs the method of transpose convolution, using transpose convolution operations to increase the image dimensions. Transpose convolution, in contrast to regular convolution, enables us to enlarge the size of feature maps during the decoding phase [54,55]. This approach aids in recovering lost details and spatial resolution, while effectively reprojecting abstract features from the encoder to the decoder. As a result, it provides a more precise and accurate feature representation for the final segmentation outcome.

Fourthly, we extract feature maps from each layer of the decoding stage, with dimensions of [512, 32, 32], [256, 64, 64], [128, 128, 128], and [64, 256, 256], respectively. Immediately following, we pass these feature maps through a feature extraction network module to further refine their information. Subsequently, we employ bilinear interpolation to perform upsampling on these feature maps, yielding four single-channel feature maps named a , b , c , and d , each with dimensions [1, 256, 256]. The bilinear interpolation algorithm calculates the weighted sum of the pixel values of the four nearest neighboring positions to the location in a smooth manner to obtain the pixel value at the target position [56] as illustrated in Figure 3. This method utilizes the linear relationship between neighboring pixels, enabling more accurate estimation of new pixel values during image enlargement. The calculation formula is shown as Equation (1):

$$\begin{aligned} a &= (1 - x) \cdot (1 - y), \\ b &= x \cdot (1 - y), c = (1 - x) \cdot y, d = x \cdot y, \\ P_0 &= aP_1 + bP_2 + cP_3 + dP_4 \end{aligned} \quad (1)$$

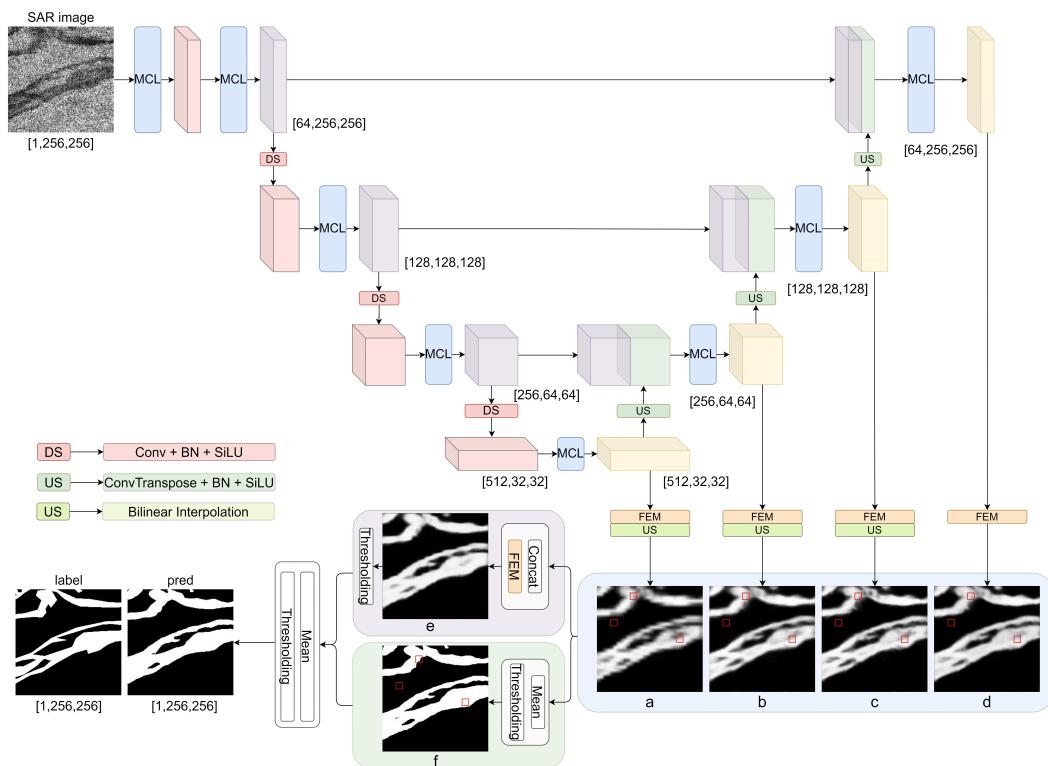


Figure 2. Overall flow of our model. We input single-channel SAR images into the model. After the U-Net encoding phase, during the decoding process, we extract the feature maps from each layer for further feature extraction. As a result, richer contextual information is captured, providing more valuable information for generating the final segmentation mask.

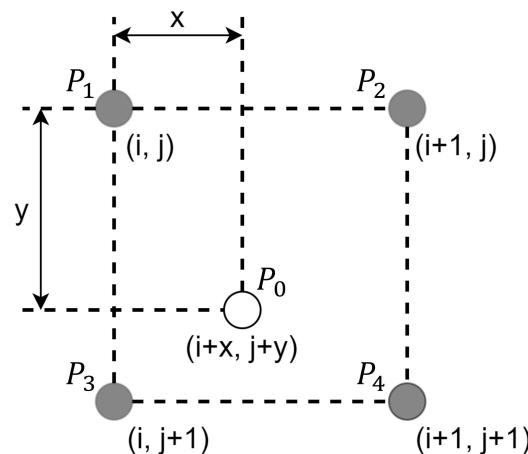


Figure 3. Bilinear interpolation diagram. Obtain the pixel value at the target position using the surrounding values of the four nearest neighboring pixels.

Fifthly, we concatenate feature maps a , b , c , and d along the channel dimension to form a composite feature map with dimensions $[4, 256, 256]$. Subsequently, this is input to the feature extraction network module again, resulting in a feature map e with dimensions $[1, 256, 256]$. Simultaneously, for feature maps a , b , c , and d , we perform per-pixel binary voting. Specifically, we average the corresponding pixels from these four feature maps and then threshold the result using a threshold of 0.5 to obtain a feature map f with dimensions $[1, 256, 256]$. This process effectively consolidates information from multiple feature maps to achieve more accurate predictions. Finally, we threshold the feature map

e with a threshold of 0.5, perform mean fusion with f , and then threshold the result again with a threshold of 0.5 to obtain the final predicted segmentation mask.

In conclusion, the main improvement of this article is the extension and novel design of the classical U-Net network architecture. By incorporating different hierarchical feature information from the decoder, the proposed network architecture obtains a wider range of information. This improvement overcomes the limitations of interplay between different levels of feature information and enables better utilization of global and local contextual information. As a result, the model exhibits enhanced information interaction and contextual awareness, effectively addressing the segmentation accuracy challenges posed by the unique imaging characteristics of SAR images.

3.2. Multi-Convolutional Layer Module

We have constructed a multi-convolutional layer (MCL) module for utilization in the encoding and decoding stages of U-Net. The aim is to extract essential feature information from the input image. This design is presented as depicted in Figure 4. “Conv” signifies a convolution operation with a kernel size of 3×3 , while “SiLU” is a novel activation function that outperforms “ReLU” [57,58], with its computational formula as shown in Equation (2):

$$f(x) = \frac{x}{1 + e^{-x}} \quad (2)$$

“BN” represents the batch normalization operation, which speeds up training convergence and enhances model stability. The input data first undergo multiple layers of convolution and a single layer of convolution operation, respectively. Subsequently, they are concatenated along the channel dimension, resulting in a more enriched feature representation. Finally, the process involves passing through another convolutional block for feature extraction and adjusting the output channel numbers. As shown in Figure 1, by using this module multiple times within the U-Net architecture, it effectively achieves multi-level feature extraction and representation of the input data.

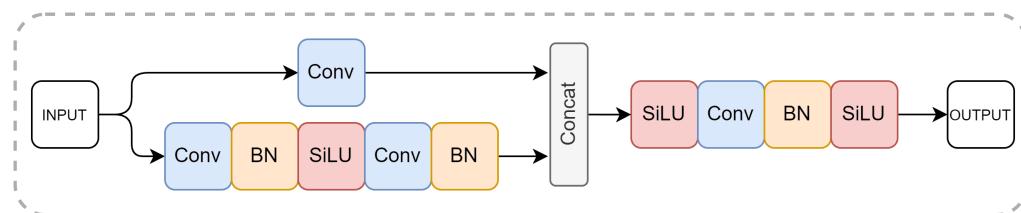


Figure 4. Multi-convolutional layer module. It is the basic unit module of the encoder and decoder in the U-Net architecture, implementing the feature extraction functionality for the input data.

3.3. Feature Extraction Module

After extracting the feature maps at each layer of the U-Net decoding stage, we designed a feature extraction module (FEM) to further extract features and adjust channel numbers. The structure of this module is illustrated in Figure 5.

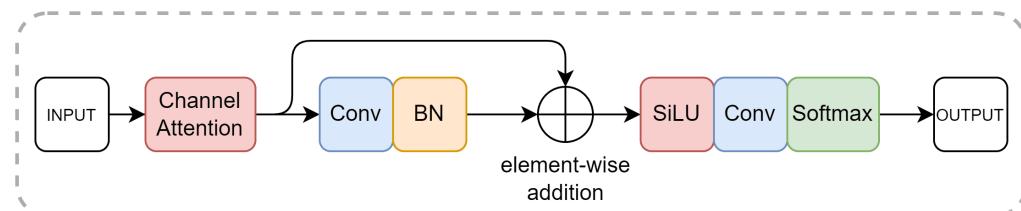


Figure 5. Feature extraction module. This module further extracts features from the feature maps obtained at each layer of the encoding stage in order to acquire more comprehensive information.

Firstly, by applying a channel attention mechanism to the input feature map, it achieves weighted modulation of the feature map along the channel dimension. This process enhances the network's focus on different channel information. Subsequently, we transmit the feature map acquired from the attention module to a convolutional block while concurrently adding the output of the convolutional block to it. Not utilizing channel-wise connections is due to the fact that, at this stage, there is already a considerable amount of contextual semantic information. By directly adding them, it allows for the re-introduction of contextual details, while also aiding in reducing the computational load. This contributes to enhancing the model's performance and efficiency. Finally, we adjust the channel count through a convolutional operation and map the output to the range [0, 1] using the Softmax function, resulting in the ultimate output.

The structure of the channel attention module is illustrated in Figure 6. We start by performing global max-pooling on the input feature map to aggregate spatial information, generating the feature F_{max} . Subsequently, F_{max} is fed into a multi-layer perceptron to obtain weights equal to the number of channels. Finally, these weights are multiplied with the input feature map as shown in Equation (3):

$$\begin{aligned} F_{max} &= GMP(input), \\ weights &= \sigma(W_1(\sigma(W_0(F_{max})))), \\ output &= weights \cdot input \end{aligned} \quad (3)$$

where σ represents the sigmoid function, $W_0 \in R^{C/2 \times C}$ and $W_1 \in R^{C \times C/2}$ are the weight coefficients of the multi-layer perceptron, and the “GMP” represents the Global Maximum Pooling.

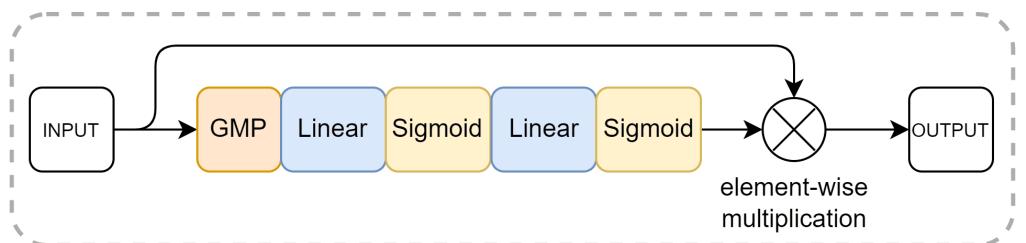


Figure 6. Channel attention module. It accomplishes weighted adjustment of the feature map along the channel dimension.

3.4. Loss Function

Using the Kullback–Leibler divergence, one can measure the difference between two separate probability distributions, p and q , for a random variable x as shown in Equation (4):

$$D_{KL}(p||q) = \sum_{i=1}^n p(x_i) \ln \frac{p(x_i)}{q(x_i)} \quad (4)$$

where n represents the number of possible values that variable x can take. When $D_{KL}(p||q)$ is smaller, it indicates that distributions p and q are closer. Expanding Equation (4) yields:

$$\begin{aligned} D_{KL}(p||q) &= \sum_{i=1}^n p(x_i) \ln p(x_i) - \sum_{i=1}^n p(x_i) \ln q(x_i) \\ &= -H(p) + H(p, q) \end{aligned} \quad (5)$$

where $H(p)$ represents the entropy of distribution p , while $H(p, q)$ represents the cross-entropy between distributions p and q . Entropy is related to the probability distribution of events; specifically, it represents the expected value of the information content of individual events within the probability distribution. In information theory, the information content of an event x is defined as

$$I(x) = -\ln(p(x)) \quad (6)$$

where $p(x)$ represents the probability of event x occurring. If the probability of event x occurring is smaller, it contains a greater amount of information when it happens. Therefore, when the probability distribution is determined, its entropy remains constant. Thus, when optimizing deep models, it is necessary to focus only on the cross-entropy between the predicted distribution and the actual distribution as shown in Equation (7):

$$H(p, q) = -\sum_{i=1}^n p(x_i) \ln q(x_i) \quad (7)$$

where n represents the number of possible values event x can take. For m random events, their cross-entropy is as shown in Equation (8):

$$J = -\sum_{i=1}^m \sum_{j=1}^n x_{i,j} \ln p(x_{i,j}) \quad (8)$$

In particular, for binary classification problems, where $n = 2$ and $x_i \in \{0, 1\}$, Equation (8) can be expanded as follows:

$$J = -\sum_{i=1}^m [x_i \ln(p(x_i)) + (1 - x_i) \ln(1 - p(x_i))] \quad (9)$$

Since the objective of oil spill detection based on SAR images is to accurately segment the oil spills from SAR images, which is a binary classification task, we use a binary cross-entropy loss function, namely:

$$L_{BCE}(p, m) = -\frac{1}{N} \sum_{i=1}^N [m_i \cdot \log(p_i) + (1 - m_i) \cdot \log(1 - p_i)] \quad (10)$$

where N represents the number of samples, m is the actual segmentation mask, and p is the predicted probability value by the model.

As shown in Figures 7–10, the feature maps extracted from the U-Net encoding stage undergo FEM and are denoted as a' , b' , c' , and d . Afterwards, a' , b' , and c' are upsampled to obtain the feature maps a , b , and c . We concatenate a , b , c , and d along the channel dimension and input them to FEM, obtaining e . To better integrate information from various feature maps, we introduce a joint loss function:

$$L_{union} = \sum_{x \in \{a, b, c, d, e\}} L_{BCE}(x, m) \quad (11)$$

This joint loss function effectively guides the network training, enabling better balance and fusion among different feature maps, thereby achieving more accurate segmentation results.

As depicted in the figures above, through the process of upsampling, the small-scale feature maps are expanded, thereby capturing more detailed information. Feature maps of different scales contain their own unique characteristics. This diversity enriches the coverage scope of contextual information, aiding in more accurately locating and identifying targets, thereby enhancing the capability of oil leak detection. The model training algorithm is depicted as shown in Algorithm 1, where “Net” represents our model as illustrated in Figure 2.

Algorithm 1 Training algorithm.

INPUT SAR images and masks dataset $D = \{(S_k, M_k)\}_k^K$
REPEAT
 Sample $(S_i, M_i) \sim D$;
 $a, b, c, d, e = Net(S_i)$;
 $L_{union} = \sum_{x \in \{a, b, c, d, e\}} L_{BCE}(x, m)$;
 Take a gradient descent step on $\nabla_\theta L_{union}$;
UNTIL convergence

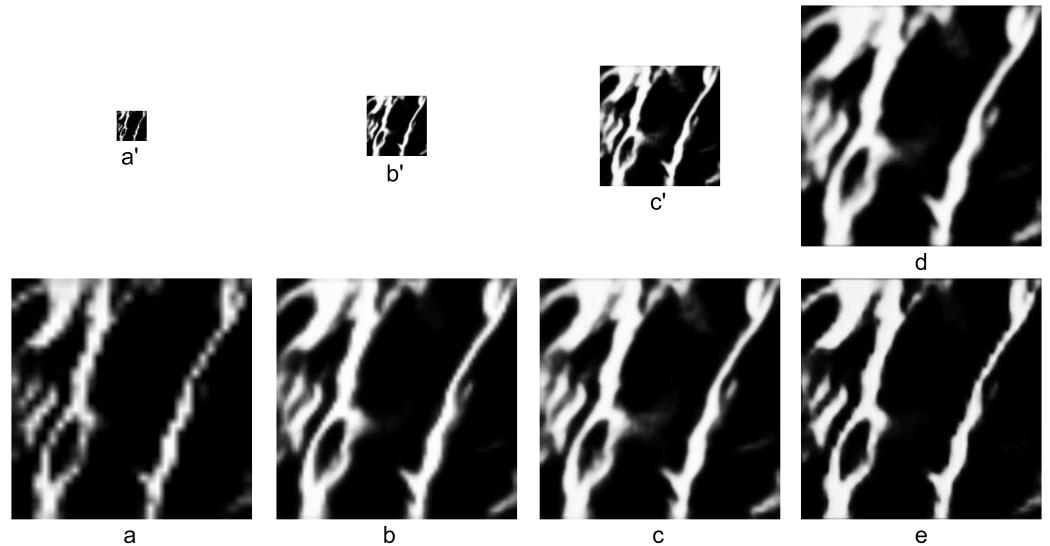


Figure 7. (a'–c') represent the feature information output by different stages of the decoder. (a–e) represent the feature information generated at different processing stages in “Net” for SAR example image 1.

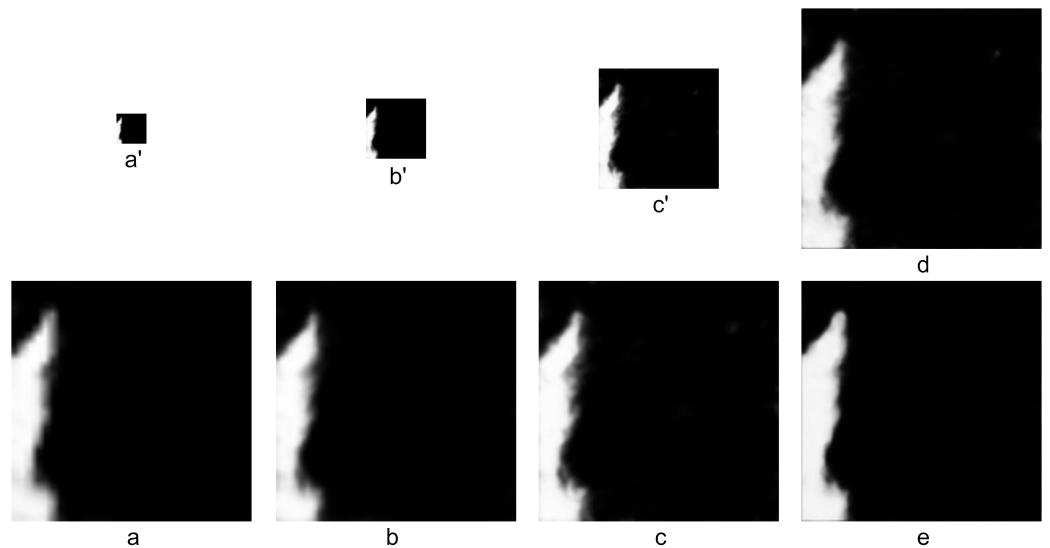


Figure 8. (a'–c') represent the feature information output by different stages of the decoder. (a–e) represent the feature information generated at different processing stages in “Net” for SAR example image 2.

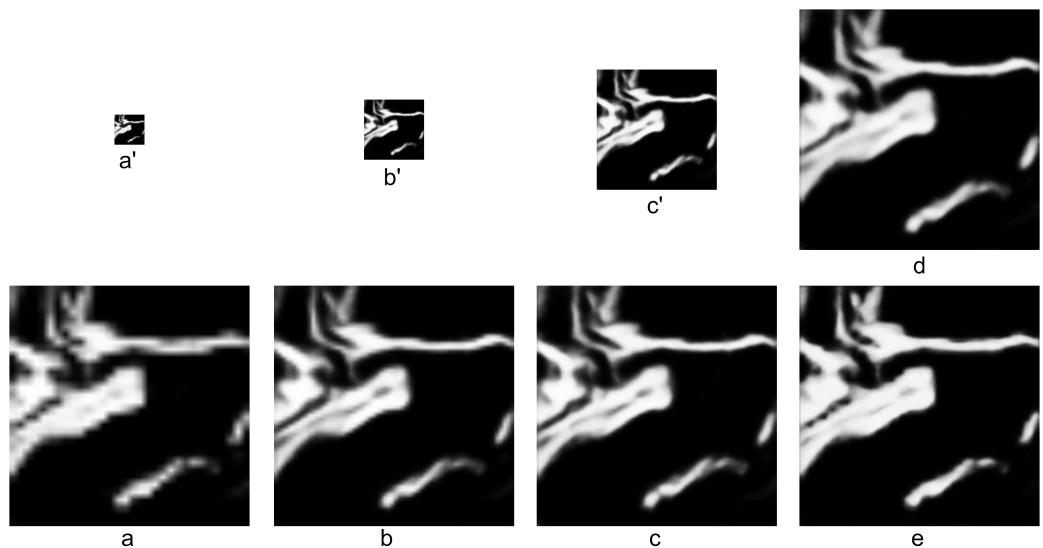


Figure 9. (a'–c') represent the feature information output by different stages of the decoder. (a–e) represent the feature information generated at different processing stages in “Net” for SAR example image 3.

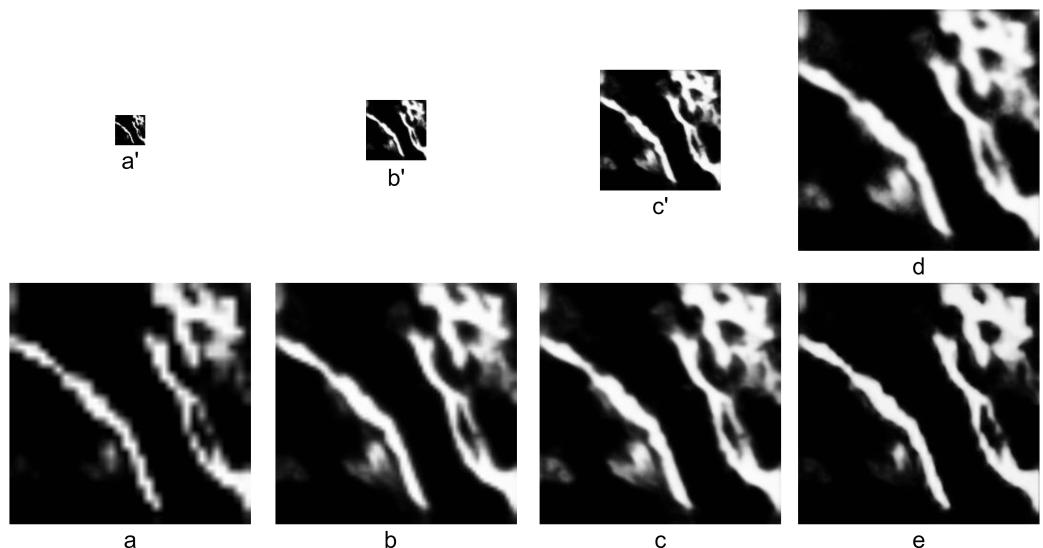


Figure 10. (a'–c') represent the feature information output by different stages of the decoder. (a–e) represent the feature information generated at different processing stages in “Net” for SAR example image 4.

4. Experimental Results and Analysis

In this chapter, we show the detailed description of the comparative experiments.

4.1. Dataset

We employ two different publicly available SAR image datasets, PALSAR and SENTINEL, to validate the effectiveness of our method for oil spill detection. The PALSAR dataset is collected by the ALOS satellite, which has L-band frequencies to enable cloud-free, and day and night land observations. The PALSAR dataset is sourced from an explosion that occurred on the Deepwater Horizon drilling platform in the Gulf of Mexico in 2010, which resulted in a massive oil spill; the spill extended approximately 160 km in length and reached a maximum width of around 72 km. The oil spill images were captured between May 2010 and August 2010, totaling 3101 training set images, each with a size of 256×256 pixels, along with their corresponding masks. Additionally, there are 776 test

set images of the same size, also accompanied by their respective masks. The PALSAR dataset is collected by Sentinel 1A satellite, which is equipped with a C-band SAR sensor and provides uninterrupted observations of the Persian Gulf region. The spatial resolution of the Sentinel 1A is 5 m in range and 20 m in azimuth. The SENTINEL dataset consists of 3345 training set images, each with a size of 256×256 pixels, along with their corresponding masks. Additionally, there are 839 test set images of the same size, also accompanied by their respective masks.

4.2. Experimental Results

We will use a range of common image semantic segmentation evaluation metrics, such as Dice score, HD95, precision, and accuracy. They respectively measure the degree of overlap in the segmentation results, the consistency of segmentation boundaries, the proportion of correctly classified pixels, and the overall segmentation accuracy. Through these evaluation metrics, we can more comprehensively assess the performance of different methods for oil spill detection based on the SAR images. In the comparative experiments, we will use the same SAR dataset to compare methods [30–37] with the approach shown in Figure 2. In the ablation experiments, we will vary certain modules of our model and validate their effectiveness once again.

As shown in Table 1, our method achieves excellent performance on two datasets, reaching the optimal levels across multiple evaluation metrics. This indicates that our approach possesses remarkable robustness and is capable of effectively accomplishing oil spill detection tasks on SAR images collected from various devices. Additionally, Figure 11 presents the masked results of the segmentation applied to eight sets of SAR images from PALSAR and SENTINEL using different methods. From these figures, the differences in segmentation outcomes among the various methods can be clearly observed.

In order to find the optimal network architecture, we take a series of steps. First, we threshold the generated images D, E, and F, and compare the computed evaluation metrics with the current results. Next, we attempt different shortcut fusion methods for the MCL and FEM modules to validate their effectiveness. Then, we adjust the height of the U-Net framework and compare the evaluation metric results. Finally, we also remove the channel attention mechanism from the FEM module to explore its impact on network performance. The experimental results are shown in Tables 2–5.

Table 1. Comparison experiment results. Bold indicates the best result, while underline indicates the second-best result. ↑ represents the higher, the better, while ↓ represents the lower, the better.

Method	Dice↑		HD95↓		Precision↑		Accuracy↑	
	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL
FCN [30]	0.723	0.741	9.95	11.17	0.790	0.655	0.897	0.827
SegNet [31]	0.667	0.664	<u>8.22</u>	<u>10.76</u>	0.605	0.519	0.899	0.803
U-Net [32]	0.716	0.733	8.79	11.05	0.636	0.709	0.911	0.807
Unet++ [33]	0.725	0.728	9.39	11.85	0.679	0.666	0.907	0.807
R2UNet [34]	<u>0.731</u>	0.709	9.05	11.02	0.694	0.640	0.909	0.787
AttU-Net [35]	0.713	0.724	8.75	11.18	0.621	<u>0.739</u>	0.912	0.790
Transunet [36]	0.721	0.732	9.04	11.94	0.593	0.690	<u>0.915</u>	0.811
Swin-Unet [37]	0.716	<u>0.755</u>	8.86	11.29	0.609	0.663	0.912	<u>0.830</u>
Ours	0.784	0.815	8.21	9.99	<u>0.729</u>	0.769	0.930	0.879

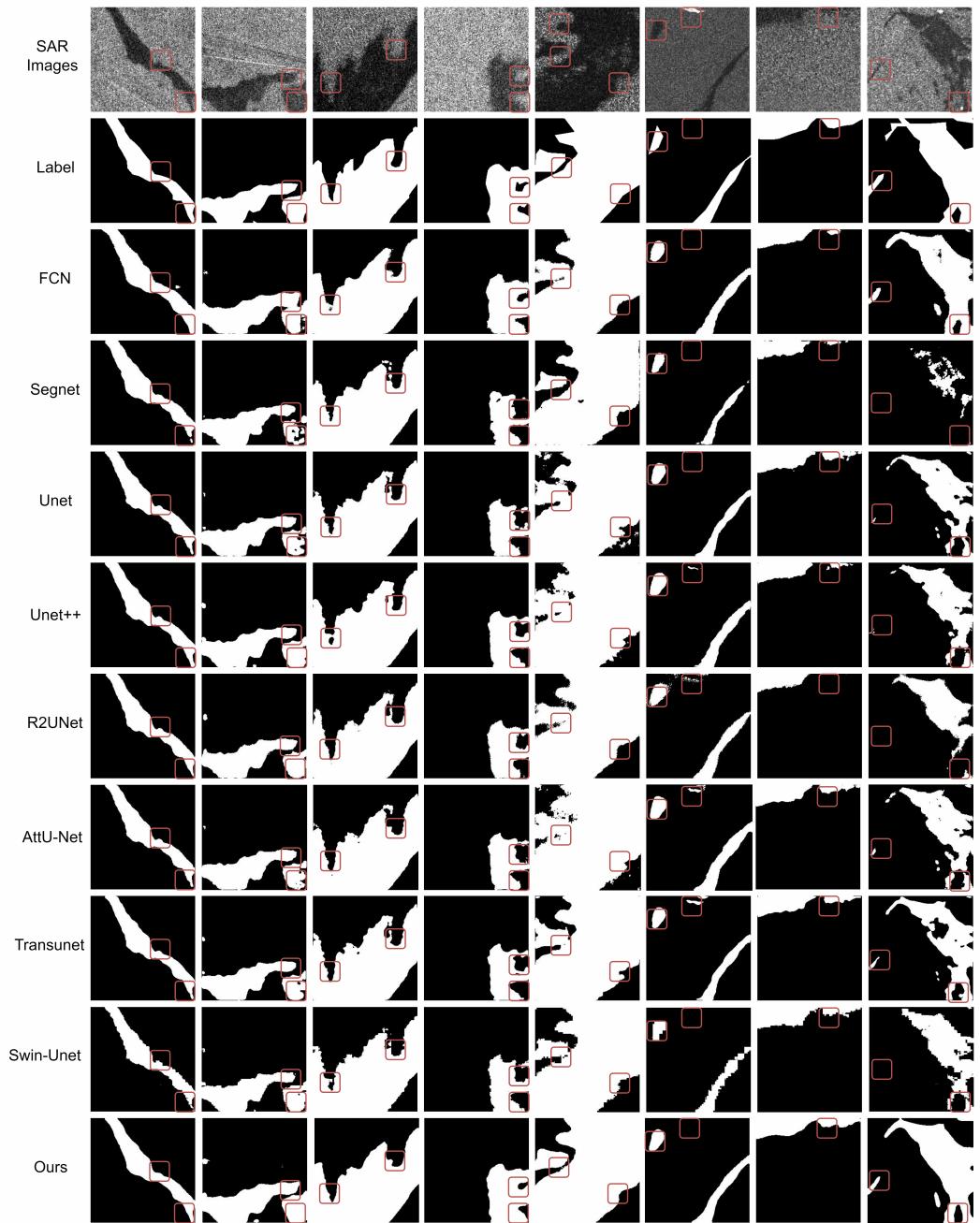


Figure 11. The visual comparison of segmentation results from different algorithm models on the same SAR image. These red boxes highlight the detailed comparison of segmentation masks generated from SAR images using different algorithms.

By conducting experiments and comparisons on different architectures, we comprehensively consider factors such as evaluation metric results and model parameter quantity. We decide to use “pred” as the final segmentation result. We apply the concatenation shortcut fusion method to the MCL module, and utilize the addition shortcut fusion method with an added channel attention mechanism for the FEM module. Simultaneously, the U-Net architecture’s height is set to four. This leads us to a network structure as shown in Figure 2 that excels in performance while maintaining a relatively efficient model complexity.

Table 2. Comparison of different mask results. Bold indicates the best result, while underline indicates the second-best result. ↑ represents the higher, the better, while ↓ represents the lower, the better.

Mask	Dice↑		HD95↓		Precision↑		Accuracy↑	
	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL
<i>d</i>	0.772	0.807	8.76	10.19	0.786	0.753	0.921	0.875
<i>e</i>	0.780	0.810	8.26	10.13	0.714	<u>0.762</u>	0.929	0.875
<i>f</i>	<u>0.781</u>	<u>0.811</u>	8.19	9.97	0.714	0.747	0.930	0.879
<i>pred</i>	0.784	0.815	<u>8.21</u>	<u>9.99</u>	<u>0.729</u>	0.769	0.930	0.879

Table 3. Comparison of the effectiveness of concatenation (C) and addition (A) as two distinct shortcut fusion methods in MCL and FEM modules. Bold indicates the best result. ↑ represents the higher, the better, while ↓ represents the lower, the better.

MCL	FEM	Dice↑		HD95↓		Precision↑		Accuracy↑	
		C	A	C	A	PALSAR	SENTINEL	PALSAR	SENTINEL
✓	✓			0.772	0.792	8.36	10.41	0.712	0.744
✓		✓		0.784	0.815	8.21	9.99	0.729	0.769
✓	✓			0.771	0.801	8.43	10.35	0.724	0.764
✓	✓			0.767	0.797	8.69	10.13	0.757	0.719

Table 4. Comparison of different height of our U-Net architecture. Bold indicates the best result. ↑ represents the higher, the better, while ↓ represents the lower, the better.

Height	Params/M	Dice↑		HD95↓		Precision↑		Accuracy↑	
		PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL
<i>three</i>	7.06	0.765	0.767	8.63	10.57	0.725	0.721	0.921	0.841
<i>four</i>	28.90	0.784	0.815	8.21	9.99	0.729	0.769	0.930	0.879
<i>five</i>	116.23	0.782	0.813	8.43	10.08	0.749	0.758	0.927	0.877

Table 5. Effect of channel attention mechanism (CAM). Bold indicates the best result. ↑ represents the higher, the better, while ↓ represents the lower, the better.

CAM	Dice↑		HD95↓		Precision↑		Accuracy↑	
	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL	PALSAR	SENTINEL
✗	0.738	0.758	8.81	10.82	0.715	0.740	0.913	0.830
✓	0.784	0.815	8.21	9.99	0.729	0.769	0.930	0.879

Experimental results demonstrate that the segmentation mask “pred” obtained through further voting achieves optimal performance due to the acquisition of additional information. This further emphasizes the superiority of our approach. Meanwhile, the utilization of the concatenation fusion method in the architecture of U-Net preserves more contextual information. On the other hand, employing the addition fusion method when processing feature maps is justified by the substantial information already present in these feature maps. This not only alleviates computational burden but also potentially enhances model performance. Shallow network depths might constrain the model’s feature extraction capacity, whereas deeper network depths could elevate the risk of overfitting. The channel attention mechanism effectively adjusts the weights of feature channels, particularly in scenarios with numerous channels and intricate information. This mechanism optimizes the quality of feature representation.

5. Discussion

In comparative experiments, our model achieves excellent results in multiple metrics compared to existing semantic segmentation networks. This is attributed to our model framework breaking the limitations of information interaction between different stages. Existing semantic segmentation networks typically extract features independently at each stage and lack effective mechanisms for information transmission between stages. Due to the specific characteristics of SAR images, such as high noise, blurry boundaries, and lighting variations, existing segmentation methods often fail to achieve satisfactory results. However, through the feature extraction framework of cross-stage feature map connections, our model is better equipped to address these challenges and achieve higher segmentation performance. Specifically, our model integrates feature maps again at each feature extraction stage in the decoder, enabling cross-stage connections of feature maps to facilitate the fusion and interaction of information between different stages. This approach is particularly effective in addressing the segmentation challenges posed by SAR images. By utilizing these cross-stage connections of feature maps, our model is able to capture global contextual information more effectively. The feature maps from different stages can transmit and share crucial semantic information, thereby enhancing the understanding of complex scenes in the image. This global contextual awareness contributes to accurate segmentation of SAR images with features such as blurry boundaries, irregular shapes, multiple speckles, and shadows. Additionally, the cross-stage connections of feature maps facilitate information fusion and updating at different scales and levels. By combining low-level features with high-level semantic features, our model can simultaneously capture both detailed and holistic semantic information, thus improving the accuracy and robustness of segmentation. This capability is particularly crucial for addressing challenges such as high noise and oil spill regions in SAR images.

The visualization results of different segmentation algorithms are shown in Figure 11. The first row represents the SAR satellite remote sensing image to be segmented, and the second row shows its corresponding segmentation mask label. The remaining rows display the segmentation masks generated by different segmentation algorithms, with the last row representing the segmentation mask generated by the proposed model in this paper. Comparing the segmentation masks produced by different algorithms with the true labels, it can be observed from some details, such as the area enclosed by the red box, that our model has achieved better results. Our model can more accurately capture detailed features of regions with complex textures and blurry boundaries, which are difficult for traditional FCN, SegNet, and other U-Net series algorithms. It can precisely identify oil spill areas and segment the boundaries between targets and backgrounds. This is attributed to the proposed model breaking the limitations of hierarchical feature interaction in the feature extraction and segmentation process, allowing our model to better utilize global and local contextual information. Through sufficient information interaction and contextual awareness, our model achieves higher accuracy and precision.

In the ablation experiments, we further explore the architecture of the model to find the optimal network design. Firstly, we threshold the masks generated by the model at different stages and calculate their segmentation performance. The results show that the masks fused multiple times achieve the highest segmentation accuracy. Secondly, we conduct ablation experiments on the feature map connections of the MCL and FEM modules. The experimental results indicate that within the U-shaped network, using concatenation connections and addition connections when extracting different-scale feature maps yields better results. Thirdly, we conduct ablation experiments on the height of the U-shaped network and find that the model architecture with a height of four layers performs better, considering both performance and parameter quantity. Finally, we perform ablation experiments on the channel attention mechanism within the FEM module. The results showed that adding channel-wise attention in the feature extraction and fusion of multi-channel feature maps at different stages of the decoder significantly improves the extraction

capacity of relevant information, thereby enhancing the segmentation performance of the model.

6. Conclusions

Oil spill pollution has escalated into a pressing global concern. In this paper, we propose a novel Oil spill detection model for SAR image, which can blend the contextual information generated by multiple scales feature maps in the U-Net. Specifically, we design a multi-convolutional layer (MCL) module to enlarge the size of feature maps during the decoding and aid in recovering lost details and spatial resolution. We also propose a feature extraction module (FEM) to enhance the ability of the proposed model to understand information from different dimension channels. To validate the efficacy of our approach, we conduct comparative experiments on two SAR datasets. The results demonstrate that, in comparison to existing segmentation networks, our proposed model excels in segmenting SAR images and achieves remarkable accuracy in oil spill detection tasks. In future research, we intend to further explore the development of an efficient algorithmic framework based on convolutional neural networks for segmenting SAR images. This endeavor aims to further elevate the performance of our methodology in oil spill detection tasks, ultimately contributing to more effective environmental monitoring and protection efforts.

Author Contributions: Conceptualization: C.L.; methodology: Y.Y.; validation: D.C.; writing—original draft preparation: Y.Y.; writing—review and editing: W.C. and X.Y.; funding acquisition: D.C.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Major Scientific and Technological Innovation Project of Shandong Province of China (2021ZLGX05, 2020CXGC010705) and National Natural Science Foundation of China under Grant 62301174, and Guangzhou Basic and Applied Basic Research Topic (Young Doctor “Sailing” Project) under 2024A04J2081.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, D.; Wan, J.; Liu, S.; Chen, Y.; Yasir, M.; Xu, M.; Ren, P. BO-DRNet: An improved deep learning model for oil spill detection by polarimetric features from SAR images. *Remote Sens.* **2022**, *14*, 264. [[CrossRef](#)]
2. Naz, S.; Iqbal, M.F.; Mahmood, I.; Allam, M. Marine oil spill detection using synthetic aperture radar over Indian Ocean. *Mar. Pollut. Bull.* **2021**, *162*, 111921. [[CrossRef](#)] [[PubMed](#)]
3. Mera, D.; Cotos, J.M.; Varela-Pet, J.; Garcia-Pineda, O. Adaptive thresholding algorithm based on SAR images and wind data to segment oil spills along the northwest coast of the Iberian Peninsula. *Mar. Pollut. Bull.* **2012**, *64*, 2090–2096. [[CrossRef](#)] [[PubMed](#)]
4. Zeng, K.; Wang, Y. A deep convolutional neural network for oil spill detection from spaceborne SAR images. *Remote Sens.* **2020**, *12*, 1015. [[CrossRef](#)]
5. Smith, L.C.; Smith, M.; Ashcroft, P. Analysis of environmental and economic damages from British Petroleum’s Deepwater Horizon oil spill. *Albany Law Rev.* **2011**, *74*, 563–585. [[CrossRef](#)]
6. Krestenitis, M.; Orfanidis, G.; Ioannidis, K.; Avgerinakis, K.; Vrochidis, S.; Kompatsiaris, I. Oil spill identification from satellite images using deep neural networks. *Remote Sens.* **2019**, *11*, 1762. [[CrossRef](#)]
7. Fingas, M.; Brown, C. Review of oil spill remote sensing. *Mar. Pollut. Bull.* **2014**, *83*, 9–23. [[CrossRef](#)] [[PubMed](#)]
8. Fingas, M.; Brown, C.E. A Review of Oil Spill Remote Sensing. *Sensors* **2018**, *18*, 91. [[CrossRef](#)] [[PubMed](#)]
9. Vasconcelos, R.N.; Lima, A.T.C.; Lentini, C.A.; Miranda, G.V.; Mendonça, L.F.; Silva, M.A.; Porsani, M.J. Oil spill detection and mapping: A 50-year bibliometric analysis. *Remote Sens.* **2020**, *12*, 3647 [[CrossRef](#)]
10. Jafarzadeh, H.; Mahdianpari, M.; Homayouni, S.; Mohammadimanesh, F.; Dabboor, M. Oil spill detection from Synthetic Aperture Radar Earth observations: a meta-analysis and comprehensive review. *Gisci. Remote Sens.* **2021**, *58*, 1022–1051. [[CrossRef](#)]
11. Picou, J.S.; Gill, D.A.; Dyer, C.L.; Curry, E.W. Disruption and stress in an Alaskan fishing community: Initial and continuing impacts of the Exxon Valdez oil spill. *Ind. Crisis Q.* **1992**, *6*, 235–257. [[CrossRef](#)]
12. Lopes, J.M.; Lentini, C.A.; Mendonça, L.F.; Lima, A.T.; Vasconcelos, R.N.; Silva, A.X.; Porsani, M.J. Absorbed dose rate for marine biota due to the oil spilled using ICRP reference animal and Monte Carlo simulation. *Appl. Radiat. Isot.* **2022**, *188*, 110354. [[CrossRef](#)] [[PubMed](#)]
13. Li, P.; Cai, Q.; Lin, W.; Chen, B.; Zhang, B. Offshore oil spill response practices and emerging challenges. *Mar. Pollut. Bull.* **2016**, *110*, 6–27. [[CrossRef](#)] [[PubMed](#)]

14. Law, R.J.; Kelly, C. The impact of the “Sea Empress” oil spill. *Aquat. Living Resour.* **2004**, *17*, 389–394. [[CrossRef](#)]
15. Brekke, C.; Solberg, A.H. Oil spill detection by satellite remote sensing. *Remote Sens. Environ.* **2005**, *95*, 1–13. [[CrossRef](#)]
16. Topouzelis, K.N. Oil spill detection by SAR images: Dark formation detection, feature extraction and classification algorithms. *Sensors* **2008**, *8*, 6642–6659. [[CrossRef](#)] [[PubMed](#)]
17. Solberg, A.H.S. Remote sensing of ocean oil-spill pollution. *Proc. IEEE* **2012**, *100*, 2931–2945. [[CrossRef](#)]
18. Pisano, A.; De Dominicis, M.; Biamino, W.; Bignami, F.; Gherardi, S.; Colao, F.; Santoleri, R. An oceanographic survey for oil spill monitoring and model forecasting validation using remote sensing and in situ data in the Mediterranean Sea. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **2016**, *133*, 132–145. [[CrossRef](#)]
19. Lu, J. Marine oil spill detection, statistics and mapping with ERS SAR imagery in south-east Asia. *Int. J. Remote Sens.* **2003**, *24*, 3013–3032. [[CrossRef](#)]
20. Fan, J.; Zhang, F.; Zhao, D.; Wang, J. Oil spill monitoring based on SAR remote sensing imagery. *Aquat. Procedia* **2015**, *3*, 112–118. [[CrossRef](#)]
21. Fustes, D.; Cantorna, D.; Dafonte, C.; Arcay, B.; Iglesias, A.; Manteiga, M. A cloud-integrated web platform for marine monitoring using GIS and remote sensing. Application to oil spill detection through SAR images. *Future Gener. Comput. Syst.* **2014**, *34*, 155–160. [[CrossRef](#)]
22. Shi, H.; Liu, Y.; He, C.; Wang, C.; Li, Y.; Zhang, Y. Analysis of infrared polarization properties of targets with rough surfaces. *Opt. Laser Technol.* **2022**, *151*, 108069. [[CrossRef](#)]
23. Sun, Z.; Zhao, Y.; Yan, G.; Li, S. Study on the hyperspectral polarized reflection characteristics of oil slicks on sea surfaces. *Chin. Sci. Bull.* **2011**, *56*, 1596–1602. [[CrossRef](#)]
24. Alpers, W.; Holt, B.; Zeng, K. Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sens. Environ.* **2017**, *201*, 133–147. [[CrossRef](#)]
25. Ajadi, O.A.; Meyer, F.J.; Tello, M.; Ruello, G. Oil spill detection in synthetic aperture radar images using Lipschitz-regularity and multiscale techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2389–2405. [[CrossRef](#)]
26. Fiscella, B.; Giancaspro, A.; Nirchio, F.; Pavese, P.; Trivero, P. Oil spill detection using marine SAR images. *Int. J. Remote Sens.* **2000**, *21*, 3561–3566. [[CrossRef](#)]
27. Migliaccio, M.; Tranfaglia, M.; Ermakov, S.A. A physical approach for the observation of oil spills in SAR images. *IEEE J. Ocean. Eng.* **2005**, *30*, 496–507. [[CrossRef](#)]
28. Latini, D.; Del Frate, F.; Jones, C.E. Multi-frequency and polarimetric quantitative analysis of the Gulf of Mexico oil spill event comparing different SAR systems. *Remote Sens. Environ.* **2016**, *183*, 26–42. [[CrossRef](#)]
29. Liu, P.; Zhao, C.; Li, X.; He, M.; Pichel, W. Identification of ocean oil spills in SAR imagery based on fuzzy logic algorithm. *Int. J. Remote Sens.* **2010**, *31*, 4819–4833. [[CrossRef](#)]
30. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
31. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
32. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
33. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018; pp. 3–11.
34. Alom, M.Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* **2019**, *6*, 014006. [[CrossRef](#)]
35. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
36. Wang, S.; Li, L.; Zhuang, X. AttU-Net: Attention U-Net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*; Springer International Publishing: Cham, Switzerland, 2021; pp. 302–311.
37. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2022; pp. 205–218.
38. Solberg, A.S.; Solberg, R. A large-scale evaluation of features for automatic detection of oil spills in ERS SAR images. *Int. Geosci. Remote Sens. Symp.* **1996**, *3*, 1484–1486.
39. Topouzelis, K.; Psyllos, A. Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 135–143. [[CrossRef](#)]
40. Yu, F.; Sun, W.; Li, J.; Zhao, Y.; Zhang, Y.; Chen, G. An Improved Otsu Method for Oil Spill Detection from SAR Images. *Oceanologia* **2017**, *59*, 311–317. [[CrossRef](#)]
41. Singha, S.; Vespe, M.; Trieschmann, O. Automatic Synthetic Aperture Radar based oil spill detection and performance estimation via a semi-automatic operational service benchmark. *Mar. Pollut. Bull.* **2013**, *73*, 199–209. [[CrossRef](#)]

42. Hu, G.; Xiao, X. Edge detection of oil spill using SAR image. In Proceedings of the 2013 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, Chengdu, China, 21–25 July 2013; pp. 466–469.
43. Liu, P.; Li, Y.; Liu, B.; Chen, P.; Xu, J. Semi-automatic oil spill detection on X-band marine radar images using texture analysis, machine learning, and adaptive thresholding. *Remote Sens.* **2019**, *11*, 756. [[CrossRef](#)]
44. Araújo, R.T.; de Medeiros, F.N.; Costa, R.C.; Marques, R.C.; Moreira, R.B.; Silva, J.L. Locating oil spill in SAR images using wavelets and region growing. In Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Hradec Kralove, Czech Republic, 10–12 July 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1184–1193.
45. Mera, D.; Bolon-Canedo, V.; Cotos, J.M.; Alonso-Betanzos, A. On the use of feature selection to improve the detection of sea oil spills in SAR images. *Comput. Geosci.* **2017**, *100*, 166–178. [[CrossRef](#)]
46. Wan, J.; Cheng, Y. Remote sensing monitoring of Gulf of Mexico oil spill using ENVISAT ASAR images. In Proceedings of the 2013 21st International Conference on Geoinformatics, Kaifeng, China, 20–22 June 2013; pp. 1–5.
47. Singha, S.; Bellerby, T.J.; Trieschmann, O. Satellite oil spill detection using artificial neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2355–2363. [[CrossRef](#)]
48. Topouzelis, K.; Stathakis, D.; Karathanassi, V. Investigation of genetic algorithms contribution to feature selection for oil spill detection. *Int. J. Remote Sens.* **2009**, *30*, 611–625. [[CrossRef](#)]
49. Chehresa, S.; Amirkhani, A.; Rezairad, G.A.; Mosavi, M.R. Optimum features selection for oil spill detection in SAR image. *J. Indian Soc. Remote Sens.* **2016**, *44*, 775–787. [[CrossRef](#)]
50. Fan, Y.; Rui, X.; Zhang, G.; Yu, T.; Xu, X.; Poslad, S. Feature merged network for oil spill detection using SAR images. *Remote Sens.* **2021**, *13*, 3174. [[CrossRef](#)]
51. Ma, X.; Xu, J.; Wu, P.; Kong, P. Oil spill detection based on deep convolutional neural networks using polarimetric scattering information from Sentinel-1 SAR images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [[CrossRef](#)]
52. Wang, Y.; Wang, C.; Wu, H.; Chen, P. An improved Deeplabv3+ semantic segmentation algorithm with multiple loss constraints. *PLoS ONE* **2022**, *17*, e0261582. [[CrossRef](#)]
53. Li, Y.; Lyu, X.; Frery, A.C.; Ren, P. Oil spill detection with multiscale conditional adversarial networks with small-data training. *Remote Sens.* **2021**, *13*, 2378. [[CrossRef](#)]
54. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. *arXiv* **2016**, arXiv:1603.07285.
55. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
56. Gribbon, K.T.; Johnston, C.T.; Bailey, D.G. A real-time FPGA implementation of a barrel distortion correction algorithm with bilinear interpolation. In Proceedings of the Image and Vision Computing New Zealand, Palmerston North, New Zealand, 26–28 November 2003; pp. 408–413.
57. Ramachandran, P.; Zoph, B.; Le, Q.V. Swish: A self-gated activation function. *arXiv* **2017**, arXiv:1710.05941.
58. Fatima, A.; Pethe, A. NVM device-based deep inference architecture using self-gated activation functions (Swish). In *Machine Vision and Augmented Intelligence—Theory and Applications: Select Proceedings of MAI 2021*; Springer: Singapore, 2021; pp. 33–44.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.