Article

Oil Spill Identification in Radar Images Using a Soft Attention Segmentation Model

Peng Chen 1, Hui Zhou 2,*, Ying Li 3, Bingxin Liu 1 and Peng Liu 1

1 Navigation College, Dalian Maritime University, Dalian 116026, China; chenpeng@dlmu.edu.cn (P.C.); gisbingxin@dlmu.edu.cn (B.L.); liupeng@dlmu.edu.cn (P.L.)
2 School of Computer and Software, Dalian Neusoft Information University, Dalian 116023, China
3 Environmental Information Institute, Dalian Maritime University, Dalian 116026, China; yldmu@dlmu.edu.cn
* Correspondence: zhouhui@neusoft.edu.cn; Tel.: +86-411-8483-2287

Abstract: Oil spills can cause damage to the marine environment. When an oil spill occurs in the sea, it is critical to rapidly detect and respond to it. Because of their convenience and low cost, navigational radar images are commonly employed in oil spill detection. However, they are currently only used to assess whether or not there are oil spills, and the area affected is calculated with less accuracy. The main reason for this is that there have been very few studies on how to retrieve oil spill locations. Given the above problems, this article introduces a model of image segmentation based on the soft attention mechanism. First, the semantic segmentation model was established to fully integrate multi-scale features. It takes the target detection model based on the feature pyramid network as the backbone model, including high-level semantic information and low-level location information. The channel attention method was then used for each of the feature layers of the model to calculate the weight relationship between channels to boost the model's expressive ability for extracting oil spill features. Simultaneously, a multi-task loss function was used. Finally, the public dataset of oil spills on the sea surface was used for detection. The experimental results show that the proposed method improves the segmentation accuracy of the oil spill region. At the same time, compared with segmentation models, such as PSPNet, DeepLab V3+, and Attention U-net, the segmentation accuracy based on the pixel level improved to 95.77%, and the categorical pixel accuracy increased to 96.45%.

Keywords: oil spill identification; image segmentation; soft attention mechanism

1. Introduction

With the rapid development of the global maritime transport industry, oil spills caused by collisions have become frequent. Frequent illegal sewage discharge and pipeline ruptures have also increased the risk of oil spills in the maritime transportation environment [1]. Oil spills are a global phenomenon and a serious environmental pollution issue in both open and coastal waters [2]. The quick and effective detection of an oil spill is of great significance to maritime transportation safety, ocean fisheries, search and rescue teams, emergency response services, and the restoration of marine environments [3].

The monitoring and detection of oil slicks is the main component of oil spill emergency management decision support [4]. Traditional methods of oil spill monitoring use aerial photography or field investigations, which require large amounts of manpower and material resources and have poor timeliness [5]. In the past few decades, many studies have used remote sensing data and techniques to extract oil spill information, and machine learning algorithms have been proven to be an effective method of extracting oil spill features from remote sensing data to identify oil slicks. Remote sensing technology and machine learning algorithms have been frequently used together in the identification and monitoring of oil spill slicks [6]. Currently, SAR (Synthetic Aperture Radar) is a common remote sensing tool
that can effectively monitor oil spills, as its imaging is not constrained by sunlight, climate, or clouds, and the resolution is not impacted by flight altitude, which makes it capable of obtaining remote sensing data at any time in any weather, but the location of an oil spill needs to be pre-identified [7]. Marine radars are widely installed on ships and can obtain remote sensing data quickly and conveniently at a low cost, which makes them able to fulfill the time and space requirements of oil spill real-time monitoring [8,9].

Oil slicks on the sea surface suppress the intensity of its backscatter, which generates a dark zone on the radar image. This phenomenon can be used to identify oil spills [10,11]. To achieve automatic oil spill identification and improve identification accuracy, image segmentation methods have been used, such as thresholding, watershed algorithm, and object segmentation, using edge information in remote sensing image oil spill identification [12]. Although these methods are all trying to overcome the challenge of oil spill regions being hard to distinguish, the classification identification results are still poor. Machine learning algorithms used in oil spill image segmentation and identification can overcome the limitations of traditional methods. For example, Zhang et al. (2015) [13] proposed an improved conditional random field (CRF) method for radar image segmentation that can capture contextual information in different scale-space structures to determine the location of the oil spill edges. Its segmentation accuracy reached 89%, but the segmentation results are impacted by noise in the images. Sun et al. [14] combined multiple random forest classifiers with the improved CRF, and the detailed classification accuracy reached 86.9%, but the generalization capability needs to be improved. One of the main flaws of machine learning is that it only considers binary classification of the oil slick and look-alike oil slick without considering other contextual characteristics in the oil spill identification process, such as ships, oil platforms, and islands. In radar images, the existence of complicated scenes, such as ships, islands, and land, impacts the identification of oil spills [15,16].

In recent years, deep learning has been widely used in the field of computer imaging [17]. Deep learning models have powerful automatic feature extraction and automatic learning capabilities, while supporting hidden layer data abstraction and obtaining contextual features. They have also achieved breakthroughs in image segmentation. Long et al. proposed fully convolutional networks (FCNs) for semantic segmentation and achieved a relatively high accuracy [18]. In 2017, the semantic segmentation model PSPNet (Pyramid Scene Parsing) utilized a pyramid pooling module to aggregate contextual information in different regions, thereby improving its capability for obtaining global information [19]. Features on different layers generated by the pyramid are connected to achieve the integration at different scales. Features on a higher layer include more semantics and less location information, while those on a lower layer correspond to location information. Therefore, this model combined multi-scale features to improve image segmentation performance. Attention U-net (2018) adopted an attention-based gated structure that achieved the attention mechanism through layer-by-layer supervision [20]. DANet (Dual Attention Network, 2019) adopted a self-attention mechanism to capture feature dependencies in spatial and channel dimensions, respectively [21]. These semantic segmentation models have been widely used in medical image analysis (such as tumor boundary extraction and tissue volume measurement) and have achieved great results. In the field of oil spill image detection, Chen et al. (2020) adopted the DeepLab V3 segmentation model to monitor sea surface oil spill areas [22].

To improve the accuracy of boundary segmentation of the oil spill area, this study proposed a segmentation model based on the soft attention mechanism for radar image segmentation. The model is a segmentation model based on the feature pyramid network (FPN), which introduces channel domain soft attention by assigning weight to each channel to represent the interdependency between the channel and the information of oil slick dark region. At the same time, this model adopts the optimized multitask loss function and uses the pixel-level segmentation scoring function as an indicator to evaluate the quality of the segmentation region, which is conducive to accurately segmenting the oil spill region on
the sea surface. This study used the X-band marine radar as the research subject to verify model validity.

2. Image Segmentation Model Based on the Soft Attention Mechanism

2.1. Segmentation Model Based on FPN Object Detection

The presence of speckle noise and uneven intensity are common issues in the oil spill area in remote sensing images [12]. Many dark regions in marine radar images are classified as oil spill areas, which makes the identification of oil spill areas difficult [23]. In order to achieve better results, the oil spill segmentation model divides the segmentation process into two steps: detection and segmentation. First determining the region of interest (ROI), and then segmenting this sub-region. This study introduces FPN into the image segmentation model as the backbone network, as shown in Figure 1. Multi-layer feature integration was achieved as the input image generated multi-scale features and aggregated them through the FPN [24]. Features on higher layers included more semantic information, while features on lower layers corresponded to location information. Then, up-sampling was taken from different layers and the results were aggregated to achieve the segmentation of oil spill images.

![Figure 1](image.png)

**Figure 1.** Multi-scale feature integration through the FPN model. Features are extracted from each layer, and a map is formed through up-sampling and horizontal connection operations.

Specifically, convolutional neural networks were first used to generate the feature maps C1, C2, C3, C4, and each layer was up-sampling using the nearest neighbor method by 2 times. The up-sampling mapping and bottom-up mapping on the same layer both went through a $1 \times 1 \times 512$ convolution kernel to be aggregated using element addition. After each layer was merged, a $3 \times 3 \times 256$ convolution was added. The final mapping set was P1, P2, P3, P4. Three rectangular boxes ($32 \times 32$ pixels, $64 \times 64$ pixels, $128 \times 128$ pixels) with different pixel areas are allocated on the P1, P2, P3, P4 layers. Simultaneously, multiple aspect ratios (1:2, 1:1 and 2:1) are used. These rectangular boxes are called anchor boxes. On each mapping layer, sliding the window with anchor boxes as a fixed area generates a large number of candidate frames, called proposals. Then, the Intersection over Union (IoU) of the proposals and the actual boundary box (ground truth) ratio was calculated. Proposals with an IoU greater than the threshold or those with the maximum IoU were kept as the ROI. By reducing the loss between the ROI and ground truth, location and classification were achieved.

When the object detection model performed positioning and classification, each convolutional layer included convolution and pooling processing, which is “down-sampling” where the pixel information of the image lessened and the features were extracted conducive to object detection [25]. However, the lessening of pixel information could also lead to inaccurate positioning of the object detection bounding box in some cases if there is noise and other dark regions. Object segmentation expands the down-sampled image after the detection and positioning process to the original size, and the output image is the same size as the original image and includes annotated information indicating the
possible classification of each pixel. Compared with object detection, the supplementary segmentation model can accurately segment the oil slick edge, as shown in Figure 2.

**Figure 2.** General architecture of the oil spill detection and segmentation model.

First, the convolutional network was shared with the models for localization and classification, and the feature map layers of the image were extracted. Then, the corresponding mapping layers of the ROI $iconv(m, n)$ were calculated. Up-sampling occurred using the transposed convolution to obtain the output matrix $deconv(m', n')$ sample $2^S$ times from $iconv(m, n)$ to $deconv(m', n')$, where $S$ is the number of layers. Then,

$$
m' = (m - 1) \times \text{stride} + \text{kernelsize} - 2 \times \text{padding}$$
$$n' = (n - 1) \times \text{stride} + \text{kernelsize} - 2 \times \text{padding}$$

(1)

$kernelsize$ was the size of the convolution, the zero-padding parameter $padding$ was set to 1, and the step size to $\text{stride} = 2^S$. The $deconv(m', n')$ obtained from different mapping layers using the transposed convolution and upsampling was the same size as the original image. Take subgraph of ROI mapped to the bottom layer P4 as an example, where the original image size is $224 \times 224$ pixels. Following FPN convolution, a map with a size of $14 \times 14 \times 256$ was formed. Transposed convolution and upsampling are presented in Figure 3. Here, let $S = 4$. After transposed convolution calculation, 0.5 was used as a threshold for binarization to generate a mask for the segmentation of the background and foreground. The segmentation result image was obtained by aggregating the sampling image from different layers.

**Figure 3.** Segmentation model based on FPN model detection.
2.2. Introducing the Soft Attention Mechanism

To improve the feature expression of the image, a soft attention mechanism was used for mapping layers of different scales in the backbone model FPN to capture the feature interdependencies between different channel maps and calculate the weighted value of all channel maps. The feature weight vector $\mathbf{w}$ explicitly modeled the interdependency between feature channels through learning.

To compute the channel attention efficiently, we squeeze the spatial dimension of the input feature map. For aggregating spatial information, average-pooling and max-pooling have been commonly adopted to compute spatial statistics. First, any one $H \times W \times C$ feature layer was input into the features map $\mathbf{P}$ (Shown in Figure 4), where $H$ and $W$ determine the size of the feature layer $\mathbf{P}$, and $C$ is the number of channels of the feature layer $\mathbf{P}$. For each channel, spatial global average pooling $\text{AvgPool}$ and maximum pooling $\text{MaxPool}$ at the size of $H \times W$ were undertaken to obtain two-channel description line vectors $\mathbf{P}_{\text{avg}}$ and $\mathbf{P}_{\text{max}}$ at a size of $1 \times 1 \times C$. Then, in order to better use the two spatial statistics rather than using them separately, the two fully connected layers (TFC) were shared and the ReLU activation function were used to fit complex dependencies between channels. Finally, the two-channel description line vectors were element-wise added to obtain the feature weight vector $\mathbf{w}$ sized $1 \times 1 \times C$ through the Sigmoid activation function, and the output weight value is controlled in the range of $[0, 1]$, as shown in Figure 5. The dot product of the original feature layer $\mathbf{P}$ ($H \times W \times C$) and and the feature weight vector $\mathbf{w}$ was determined to obtain the feature layers with different levels of importance for different channels. The layers with channel weights are then merged at each layer in the FPN manner, that is, a new feature map layer is obtained, as shown in Figure 4.

$$\mathbf{w} = \text{sigmoid}(\text{TFC}(\mathbf{P}_{\text{avg}}) + \text{TFC}(\mathbf{P}_{\text{max}}))$$

$$\mathbf{P}_{\text{avg}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \text{AvgPool}(\mathbf{P})$$

$$\mathbf{P}_{\text{max}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \text{MaxPool}(\mathbf{P})$$

Figure 4. Introducing the soft attention mechanism to FPN. The feature weight vector was used in each the convolution layer, and obtained different levels of importance for different channels.
Figure 5. The Soft Attention Mechanism, $H \times W \times C$ is the features map ($P$), the blue block represents global average pooling, the orange block represents global maximum pooling, and the colored block represents the feature weight vector.

2.3. Multitask Loss Function

There are three parts in this model: classification, positioning, and segmentation, so the multitask loss function of this model ($L$) includes these three parts.

$$L = \sum_{i} L_{\text{cls}}(p_i, u_i) + \alpha \cdot L_{\text{mask}} + \beta \cdot L_{\text{IoU}}$$

(3)

where $\alpha$ and $\beta$ are weight adjustment factors, $i$ represents the $i$th selected ROI. In the training process, the classification loss function included multi-class classification loss and pixel classification loss, which were supervised by a cross-entropy loss function.

The multi-class classification loss was $L_{\text{cls}}(p_i, u_i) = -\log(p_{u_i})$, every selected ROI had a probability distribution of $p_i$, and if the calculated candidate box had a positive label, then $u_i = 1$; if it had a negative label, then $u_i = 0$.

$L_{\text{mask}}$ segmentation loss function is also a binary cross entropy classification loss function based on pixel calculation that can tell whether a pixel is in the foreground or background. Every segmentation result contained $N_{\text{pixel}}$ pixels; therefore, $L_{\text{mask}}$ is the mean of binary cross-entropy loss for all pixels in the segmentation result of a selected ROI, which is as follows:

$$L_{\text{mask}} = \frac{1}{N_{\text{pixel}}} \sum_{j=1}^{N_{\text{pixel}}} - (y_j \times \log p_j + (1 - y_j) \times \log p_j)$$

(4)

$L_{\text{IoU}}$ is a location loss function based on IoU. IoU is the intersection over union of the predicted box and ground truth, which reflects the level of coincidence of the predicted box and ground truth. The higher the coincidence, the higher the IoU value. Therefore, $1 - \text{IoU}$ can be used as the loss function.

$$L_{\text{IoU}} = \sum_{i} p_i [1 - \text{IoU}]^2 + \lambda \sum_{i} (1 - p_i) [1 - \text{IoU}]^2$$

(5)

where $\lambda$ is the penalty factor and is the sensitivity. Given the coordinates of the ground truth are $\text{gt}$, and the calculated predicted box coordinates are $\text{pb}$, IoU can be obtained through calculations.

$$\text{pb} = (x_{p_{\text{min}}}, x_{p_{\text{max}}}, y_{p_{\text{min}}}, y_{p_{\text{max}}})$$

$$\text{gt} = (x_{\text{min}}, x_{\text{max}}, y_{\text{min}}, y_{\text{max}})$$

$$A_p = (x_{p_{\text{max}}} - x_{p_{\text{min}}}) \times (y_{p_{\text{max}}} - y_{p_{\text{min}}})$$

$$A_g = (x_{\text{max}} - x_{\text{min}}) \times (y_{\text{max}} - y_{\text{min}})$$

$$I_{pg} = \begin{cases} (x_2 - x_1) \times (y_2 - y_1), & \text{if } x_2 > x_1, y_2 > y_1 \\ 0, & \text{otherwise} \end{cases}$$

(7)

$$U_{pg} = A_p + A_g - I_{pg}$$

$$\text{IoU} = \frac{I_{pg}}{U_{pg}}$$

(8)
where \( x'_1 = \max(x_{\text{min}}^p, x_{\text{min}}^g), x'_2 = \min(x_{\text{min}}^p, x_{\text{min}}^g), y'_1 = \max(y_{\text{min}}^p, y_{\text{min}}^g), y'_2 = \min(y_{\text{min}}^p, y_{\text{min}}^g) \), \( I_{pg} \) is the intersection of the predicted box and ground truth, and \( U_{pg} \) is the union of the predicted box and ground truth.

### 3. Experimental Analysis

#### 3.1. Dataset

The experiment first used the SAR image dataset pre-trained model provided by the four research institutes of the Russian Academy of Sciences, the University of Hamburg, the National University of Singapore, and the National Central University on tropical and subtropical marine ERS-2 SAR images for transfer learning [26]. Then use the marine radar oil spill dataset. This dataset uses the X-band marine radar installed on the Yukun Wheel from Sperry Marine [27], and the Figure 6 shows the voyage path for this data acquisition. The range resolution of marine radar was 3.75 m and its azimuth resolution was 0.1°. The oil spill radar image scan range detection radius was set to 1.389 km. Other major parameters are shown in Table 1. Figure 7 shows an example of a radar image with a scan radius of 0.75 nautical miles at 23:19 on 21 July 2010. Figure 8 is the converted X-band marine radar image. Data enhancement, including translation, flipping, and rotation, was carried out in this experiment. Image data were divided into background and oil spill areas.

| Operation frequency | 9.41 GHz |
|---------------------|----------|
| Antenna type        | Slotted waveguide antenna |
| Range               | ~0.1–5.0 km |
| Detection angle     | Horizontal 360° |
|                     | Vertical ±10° |
| Impulse repetition frequency | 3000 Hz/1800 Hz/785 Hz |
| Impulse width       | 50 ns/250 ns/750 ns |

Figure 6. Navigation route of ship “YUKUN”. Oil spill sampling points are marked in yellow.
3.2. Experiment Process

The experiment platform was Ubuntu16.0, the GPU was NVIDIA Tesla V100, and the development platform was Paddle X. During the experiment, the empirical learning rate of the semantic model was 0.00001, the batch_size was 24, and the dataset was randomly arranged in each iteration. Its loss function is shown in Figure 9 below.

IoU, the intersection over union of the identification box and ground truth, was used as the evaluation indicator. In the oil spill identification, we are not only concerned with whether the oil spill identification was correct, but also how accurate the oil spill regions edge was. Therefore, in the evaluation of a semantic segmentation model, the quality of the segmentation results was also very important.

This study used pixel-level IoU $S_{mask \_IoU}$ and categorical pixel accuracy $S_{CPA}$ as the evaluation indicators for segmentation tasks. $S_{mask \_IoU}$ is the task score for pixel-level semantic segmentation, which is the IoU of the predicted and true semantic segmentation results. $N_{ij}$ represents the number of pixels on the oil slick predicted as the oil slick, $N_{ji}$ represents the number of pixels on the oil slick not predicted as the oil slick, and $N_{ji}$ represents the number of pixels not on the oil slick predicted as the oil slick. $S_{CPA}$ is the
categorical pixel accuracy, which is the ratio of correctly predicted pixels on the oil slick to the total number of predicted pixels.

\[
S_{\text{mask}_{\text{IoU}}} = \frac{N_{ii}}{N_{ij} + N_{ji} - N_{ii}} \quad (10)
\]

\[
S_{\text{CPA}} = \frac{N_{ii}}{N_{ii} + N_{ji}} \quad (11)
\]

The sea surface oil spill segmentation results are shown in Figure 10. Because of the introduction of the soft attention mechanism to calculate the interdependency between the
dark region and the channel information, the performance was great. The accuracy of the segmentation tasks reached 95.77%, and the categorical pixel accuracy reached 96.45%, as shown in Table 2.

Table 2. Comparison of optimization after introducing different improvement modules.

| FCN | FPN | Soft Attention | Multi-Task Loss | S_mask_IoU (%) | S_CPA (%) |
|-----|-----|----------------|-----------------|----------------|-----------|
| √   |     |                |                 | 75.12          | 79.49     |
| √   | √   |                |                 | 88.70          | 92.14     |
| √   | √   | √              |                 | 95.77          | 96.45     |

Figure 10. Sea surface oil spill segmentation results. (a) The image, (b) the ground truth, (c) method proposed in this study.

VGG19, ResNet50, and FPN were used as the backbone network, respectively, and FCN was used as the detector to detect the oil spill on the sea surface. S_mask_IoU and S_CPA were used as the evaluation indicator. As shown in Table 3, after the introduction of the soft attention mechanism to calculate channel weight with different backbone networks, the average segmentation accuracy increased by 6.04% and the average categorical pixel accuracy increased by 4.79%.
Table 3. Comparison of oil spill detection models $S_{mask_{IoU}}$ and $S_{CPA}$.

| Backbone          | Detector | $S_{mask_{IoU}}$ (%) | $S_{CPA}$ (%) |
|-------------------|----------|----------------------|---------------|
| VGG19             | FCN      | 75.12                | 79.49         |
| VGG19 + soft attention | FCN      | 80.95                | 85.75         |
| ResNet50          | FCN      | 82.85                | 86.06         |
| ResNet50 + soft attention | FCN      | 87.43                | 89.86         |
| FPN               | FCN      | 88.70                | 92.14         |
| FPN + soft attention | FCN      | 95.77                | 96.45         |

It can be seen from the comparison of the segmentation evaluation indicators of the detection models that the accuracy of the FCN model with VGG as the backbone was 75.12%. After the introduction of the soft attention mechanism, $S_{mask_{IoU}}$ increased by 5.83% and $S_{CPA}$ increased by 6.26%. With the introduction of the soft attention mechanism, when the FCN segmentation model was switched to FPN, the detection performance increased further, with $S_{mask_{IoU}}$ and $S_{CPA}$ reaching 95.77% and 96.45% respectively. This indicates that on the basis of the FPN backbone model integrated with multi-scale semantic features, using the soft attention mechanism to calculate the weight of each channel and establishing the interdependency between feature channels can effectively improve the semantic segmentation model and improve the accuracy of oil spill detection. The oil spill segmentation results with different backbone networks after the introduction of the soft attention mechanism are shown in Figure 11.

![Figure 11](image-url)
3.3. Comparison with Other Models

Comparing the model purposed in this study with other image segmentation models using the dataset built in this study with $S_{\text{mask}_{\text{IoU}}}$ and $S_{\text{CPA}}$ as evaluation indicators, the results are shown in Table 4. The $S_{\text{mask}_{\text{IoU}}}$ value and $S_{\text{CPA}}$ values of the PSPNet model were 87.06% and 91.38%, respectively. This model also adopted the pyramid pooling module to aggregate contextual information in different regions combined with multi-scale features to improve image segmentation performance. The $S_{\text{mask}_{\text{IoU}}}$ value and $S_{\text{CPA}}$ values of the DeepLab V3+ segmentation model were 92.18% and 94.90%, respectively, and it mainly improved the segmentation accuracy on the edge by introducing a CRF, thereby increasing detection accuracy. The $S_{\text{mask}_{\text{IoU}}}$ and $S_{\text{CPA}}$ values of the attention U-net semantic segmentation model were 93.32% and 94.72%, respectively. This model also introduced the attention mechanism, and it was based on an attention-gate structure, which also led to high accuracy. Compared with other relatively new models, the segmentation model proposed in this study for oil spill detection adopted the FPN model integrated with multi-layer semantic features, and also introduced the channel attention mechanism to strengthen the different weight for each channel, which further improved detection performance. The difference of $S_{\text{mask}_{\text{IoU}}}$ between this model and PSPNet, DeepLab V3+, Attention U-net is 8.71%, 3.59% and 2.45%, the difference of $S_{\text{CPA}}$ is 5.07%, 1.55% and 1.73%, yielding an $S_{\text{mask}_{\text{IoU}}}$ value of 95.77% and an $S_{\text{CPA}}$ value of 96.45%.

Table 4. Comparison of $S_{\text{mask}_{\text{IoU}}}$ and $S_{\text{CPA}}$ from different segmentation models.

| Model                      | $S_{\text{mask}_{\text{IoU}}}$ (%) | $S_{\text{CPA}}$ (%) |
|----------------------------|-----------------------------------|----------------------|
| PSPNet [19]                | 87.06                             | 91.38                |
| DeepLab V3+ [22]           | 92.18                             | 94.9                 |
| Attention U-net [20]       | 93.32                             | 94.72                |
| Model Purposed in this Study | 95.77                             | 96.45                |

After the marine radar received the wave echo signal, the original signal image data of the sea clutter information were generated. The original marine radar data often contained co-channel interference, which increased classification difficulty. Figure 12 includes semantic segmentation results of the PSPNet, DeepLab V3+, and Attention U-net object detection plus semantic analysis models and the model proposed in this study, respectively. The rectangles mark the oil spill region where segmentation was not accurate, and ovals mark oil spill regions that were missed. The PSPNet basically detected correctly, but missed some small regions compared to other models. The DeepLab V3+ model captured details at the edges, but its segmentation of oil slicks with complicated shapes was not accurate. The Attention U-net model had a fewer misses after the introduction of the attention mechanism but had some inaccuracy issues when the oil slick expanded. The model purposed in this article adopted the object detection network as its backbone network. It detected ever oil spill region firstly and then conducted semantic segmentation on each detected region, which was a semantic segmentation based on instance, thereby significantly increasing segmentation accuracy. At the same time, the soft attention mechanism was introduced during the generation of convolution feature images to calculate channel weight, providing great detailed segmentation for the object and further improving detection accuracy.
Figure 12. The segmentation results of different module experiments: (a) The image, (b) PSPNet, (c) DeepLab V3+, (d) Attention u-net, and (e) FPN + soft attention + FCN. The rectangles mark the oil spill regions where segmentation was not accurate, and ovals mark oil spill regions that were missed.

4. Conclusions

This study proposed a segmentation model based on the soft attention mechanism that adopted the object detection network based on the feature pyramid as the backbone network combined with semantic integration. Channel soft attention was introduced to assign a weight to each channel to represent the interdependency between the channel and information of the oil slick dark region, overcoming the poor classification of satellite images for oil spill monitoring and improving the capability of capturing the fine details of the object. Monitoring oil spill areas through marine radar images, the model proposed in this study achieved 95.77% accuracy for the segmentation indicator and 96.45% classification accuracy, performing excellently for remote sensing image oil spill classification. The improved model is of great significance to marine environment restoration and sea surface pollution checks. However, the network model built in this study depends heavily on a large number of annotated images, which requires considerable experienced manpower and will be affected by subjective factors. Therefore, segmentation models based on lightly supervised learning will be the focus of future research, and this is conducive to improving the feasibility of the algorithms.

Author Contributions: P.C. conceived and designed the algorithm and contributed to the manuscript and experiments; H.Z. was responsible for the construction of ship detection dataset, constructed the outline for the manuscript and made the first draft of the manuscript; Y.L. and B.L. supervised the experiments and were also responsible for the dataset; P.L. carried out oil spill detection by machine learning methods. All authors have read and agreed to the published version of the manuscript.
Funding: This research was supported by “the Fundamental Research Funds for the Central Universities”, grant number 3132022141.

Data Availability Statement: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

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

References
1. Delpeche-Ellmann, N.C.; Soomere, T. Investigating the Marine Protected Areas most at risk of current-driven pollution in the Gulf of Finland, the Baltic Sea, using a Lagrangian transport model. *Mar. Pollut. Bull.* 2013, 67, 121–129. [CrossRef] [PubMed]
2. Al-Ruzouq, R.; Gibril, M.B.A.; Shanableh, A.; Kais, A.; Hamed, O.; Al-Mansoori, S.; Khalil, M.A. Sensors, features, and machine learning for oil spill detection and monitoring: A review. *Remote Sens.* 2020, 12, 3338. [CrossRef]
3. Alves, T.M.; Kokinou, E.; Zodiatis, G.; Radhakrishnan, H.; Panagiotakis, C.; Lardner, R. Multidisciplinary oil spill modeling to protect coastal communities and the environment of the Eastern Mediterranean Sea. *Sci. Rep.* 2016, 6, 1–9. [CrossRef] [PubMed]
4. Fingas, M.; Brown, C.E. Oil spill remote sensing: A forensics approach. In *Standard Handbook Oil Spill Environmental Forensics*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 961–981.
5. Fingas, M.; Brown, C.E. A review of oil spill remote sensing. *Sensors* 2018, 18, 91. [CrossRef] [PubMed]
6. Ober, H.K. *Effects of Oil Spills on Marine and Coastal Wildlife*; EDIS; UF/IFAS North Florida Research and Education Center: Quincy, FL, USA, 2010.
7. Alpers, W.; Holt, B.; Zeng, K. Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sens. Environ.* 2017, 201, 133–147. [CrossRef]
8. Gangeskar, R. Automatic oil-spill detection by marine X-band radars. *Sea Technol.* 2004, 45, 40–45.
9. Liu, P.; Li, Y.; Xu, J.; Zhu, X. Adaptive enhancement of X-band marine radar imagery to detect oil spill segments. *Sensors* 2017, 17, 2349. [CrossRef] [PubMed]
10. Xu, J.; Jia, B.; Pan, X.; Li, R.; Cao, L.; Cui, C.; Wang, H.; Li, B. Hydrographic data inspection and disaster monitoring using shipborne radar small range images with electronic navigation chart. *PeerJ Comput. Sci.* 2020, 6, e290. [CrossRef] [PubMed]
11. Xu, J.; Wang, H.; Cui, C.; Zhao, B.; Li, B. Oil spill monitoring of shipborne radar image features using SVM and local adaptive threshold. *Algorithms* 2020, 13, 69. [CrossRef]
12. Ozigis, M.S.; Kaduk, J.D.; Jarvis, C.H. Mapping terrestrial oil spill impact using machine learning random forest and Landsat 8 OLI imagery: A case site within the Niger Delta region of Nigeria. *Environ. Sci. Pollut. Res.* 2019, 26, 3621–3635. [CrossRef] [PubMed]
13. Zhang, P.; Li, M.; Wu, Y.; Li, H. Hierarchical conditional random fields model for semisupervised SAR image segmentation. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 4933–4951. [CrossRef]
14. Sun, X.; Lin, X.; Shen, S.; Hu, Z. High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. *ISPRS Int. J. Geo-Inf.* 2017, 6, 245. [CrossRef]
15. Zhu, X.; Li, Y.; Feng, H.; Liu, B.; Xu, J. Oil spill detection method using X-band marine radar imagery. *J. Appl. Remote Sens.* 2015, 9, 095985. [CrossRef]
16. Xu, J.; Pan, X.; Jia, B.; Wu, X.; Liu, P.; Li, B. Oil spill detection using LBP feature and K-means clustering in shipborne radar image. *J. Mar. Sci. Eng.* 2019, 7, 65. [CrossRef]
17. Fingas, M.; Brown, C. Review of oil spill remote sensing. *Mar. Pollut. Bull.* 2014, 83, 9–23. [CrossRef] [PubMed]
18. 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.
19. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
20. Lian, S.; Luo, Z.; Zhong, Z.; Lin, X.; Su, S.; Li, S. Attention guided U-Net for accurate iris segmentation. *J. Vis. Commun. Image Represent.* 2018, 56, 296–304. [CrossRef]
21. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
22. Chen, Y. Research on maritime oil spill monitoring of multi-source remote sensing image based on deep semantic segmentation. In Proceedings of the 43rd COSPAR Scientific Assembly, Sydney, Australia, 28 January–4 February 2021 2021; Volume 43, p. 105.
23. Pelta, R.; Carmon, N.; Ben-Dor, E. A machine learning approach to detect crude oil contamination in a real scenario using hyperspectral remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 2019, 82, 101901. [CrossRef]
24. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
25. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 7405–7415. [CrossRef]
26. Topouzelis, K.N. Oil spill detection by SAR images: Dark formation detection, feature extraction and classification algorithms. *Sensors* **2008**, *8*, 6642–6659. [CrossRef] [PubMed]

27. 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]