

Offshore Oil Slicks Detection From SAR Images Through The Mask-RCNN Deep Learning Model

Emna Amri, Alexandre Benoit, Philippe Bolon, Véronique Migebielle, Bruno Conche, Georges Oppenheim

▶ **To cite this version:**

Emna Amri, Alexandre Benoit, Philippe Bolon, Véronique Migebielle, Bruno Conche, et al.. Offshore Oil Slicks Detection From SAR Images Through The Mask-RCNN Deep Learning Model. 2020 IEEE International Joint Conference on Neural Networks (IJCNN), World Congress on Computational Intelligence (WCCI), Jul 2020, Glasgow (virtual), United Kingdom. pp.8, 10.1109/IJCNN48605.2020.9206652 . hal-02945431

HAL Id: hal-02945431
https://hal.science/hal-02945431

Submitted on 8 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Offshore Oil Slicks Detection From SAR Images Through The Mask-RCNN Deep Learning Model

Amri Emna
LISTIC laboratory,USMB/Total company
emna.amri@univ-smb.fr

Benoit Alexandre
LISTIC laboratory,USMB
alexandre.benoit@univ-smb.fr

Philippe Bolon
LISTIC laboratory,USMB
philippe.bolon@univ-smb.fr

Migebielle Véronique
Total company
veronique.miegebielle@total.com

Conche Bruno
Total company
bruno.conche@total.com

Oppenheim Georges
University of Paris-Est
georges.oppenheim@gmail.com

Abstract

This paper introduces a method for offshore oil slick detection. At present, Synthetic Aperture Radar (SAR) is an image acquisition technology useful for oil slick detection in all weather conditions. It is used to carry out the detection, with notable limitations under certain conditions (surfaces, weather conditions). Manual SAR images analysis is expensive and, given the increasing amount of data collected from available sensors, automation becomes mandatory. To achieve this objective, instance object detection relying on deep neural networks is interesting to adapt to the data variability. Relying on such an approach, this article explores the capabilities of generalizing the detection of slicks on large datasets using the Mask-RCNN model. A detailed performance analysis is established in two complementary directions: (i) the impact of the SAR image characteristics(sensor, geographical areas, lookalike presence), (ii) the impact of the neural network architecture, transferred capabilities and training procedures. The main findings of this analysis show that Mask-RCNN features promising performance for pollution detection.

I. Introduction

In the offshore domain, several major challenges have been identified for the successful detection of oil slicks. The first one is related to the speed of the image acquisition process, as obtaining quick information is decisive in case of events as oil spillages to react as fast as possible. Recent satellite launches, however? are improving the ability to acquire images by providing global coverage and 24-hour acquisition capability. On the other hand, the growing amount of data generated by the satellites images is leading us to a second challenge concerning the ability to process all the acquired images in an effective and efficient manner. A third challenge deals with accurate identification of the oil slicks, successfully differentiating them from the other phenomena on the sea surface [3]. Finally, the fourth challenge is searching for precision in oil slick localization in very large images [1]. In working to solve this challenges, SAR images have proven to be reliable for the detection of oil slicks [7] by providing high resolution images in a wide range of weather conditions. Nevertheless, the detection and segmentation of offshore oil slicks are generally carried out by human image interpreters, who spend hours processing the images to find such anomalies. The task is difficult since oil slick has high variability in terms of shape and size and could be easily confused with sim-

ilar dark structures (lookalike) coming from algae, waves, etc [22]. This paper is organized as follows: Section 2 summarizes the related literature. Section 3 presents an introduction of the deep learning approach, Section 4 describes the considered SAR data collection and the performance evaluation metrics. Section 5 presents the experimental design and section 6 discussed the obtained results and draws the main conclusions and future research.

II. Relative Works to the oil slick detection

Offshore oil slick detection has been a challenge for several years. This is mainly carried out using either passive (optical/infrared sensors) or active (microwave sensors) remote sensing sensors. Among these sensors, active sensors have advantages in their ability to operate day and night, independently of the sunlight, unlike optical sensors. SAR Sensors constitute a powerful tool for detecting hydrocarbons on the sea surface due to the sensitivity of microwave signals to the surface roughness [31] [38]. Most of the electromagnetic energy is reflected by rough surfaces such as clean water disturbed by the wind. However, the oil slick locally dampens the roughness of the sea surface, thus decreasing the radar scattering. As a result, dark spots are formed which contrast with the brightness of the surrounding slick free sea [3]. Such phenomena should then allow the automatic detection of the oil slick. However, many lookalikes can occur above the sea surface and appear on radar images as dark spots, in the same way as areas covered by oil. [7] [35] proposes an interesting hierarchy of these distractors which is illustrated in Fig.1.

Furthermore, oil slicks are small targets in large images that remain scarce. Oil slicks indeed cover less than 1% of the image pixels as shown in the pixel Table 1. These values are calculated on a typical sample of data, they reveal the imbalance between the slick pixels and the sea and lookalike pixels. Such detection is therefore harder than multimedia object

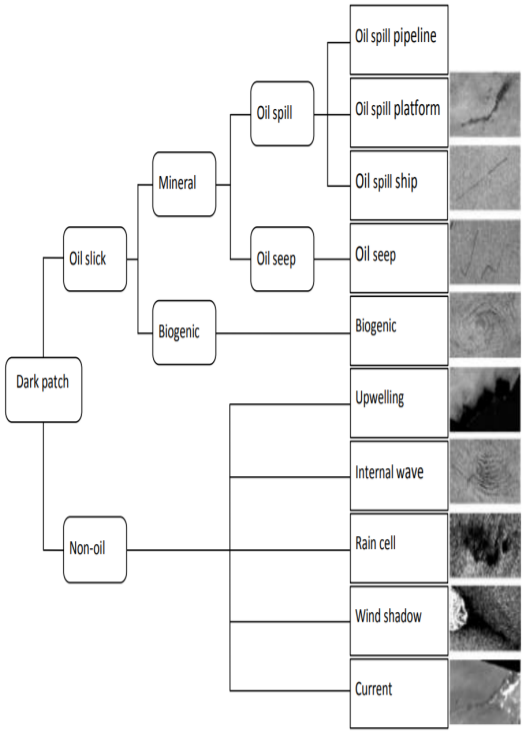


Figure 1: *Main offshore dark patches seen in SAR images, from [35].*

detection benchmarks such as Cityscapes [10] and Coco [19], which more frequently present medium and large objects compared to the input image.

Table 1: *The distribution between oil slicks and clean water samples in the considered dataset: on the left at the image cropping level, on the right at the pixel level.*

Area	Crops number	Pixels Number
Slick	8879	$\sim 2*10^7$
Sea and lookalike	423	$\sim 3*10^9$

Consequently, oil slick detection is a challenge due to numerous factors, such as the little information associated with them, the possibility of confusion with the background, the higher accuracy requirement for location

and the large size of the image [1]. The state of the art regarding oil slick detection can be summarized in Table 2. A distinction is made between semi-automatic and fully automatic approaches, which are reviewed in the following.

Table 2: *Survey on oil slick detection approaches.*

Conventional approach	Semi Automatic Partial integration of deep learning
Fully end to end deep learning	Transfer learning Full Automatic

i. Semi-Automatic Approaches

Such approaches are summarized in four steps. Detection and isolation of all dark formations that are present in the image. This is accomplished mainly through thresholding and segmentation processing [36] [17]. The second step concerns the extraction of the characteristics of the dark regions, mainly their geometrical parameters as well as their physical behaviors (e.g. mean backscatter value, polarimetry) [4] [7] and contextual data (e.g. distance to ships) [30]. The third step is the classification or differentiation of the extracted values as slick (spill, seep) or lookalike. A variety of classifiers have been used, i.e. statistical approach by probability calculation, fuzzy logic, etc. [29] [6], [3]. The fourth step examines an assessment against predefined values commonly known as annotation/ground truth/label, which establishes a classification between man-made oil slicks and lookalike phenomena. These values are generally determined through phenomenological considerations (empirical research) and statistical assessments [11]. Conventional approaches exhibit poor generalization behaviors, remaining accurate only in specific configurations (sensor, wind speed, frequency band, etc.). The main issues are related to the confusion between slick and lookalikes and the classification between oil slicks types.

To overcome these limitations, learning-based approaches such as deep neural networks are integrated into one of the above steps. This integration can be at the step of detection, features extraction, or classification of the offshore oil slicks. Several works [33] investigate the use of the capabilities of Convolutional Neural Networks (CNNs) in many steps of the classical detection process that outperform conventional approaches. For example, a text classifier based on a neural network algorithm is used for the detection of dark features by [21]. The work of [14] employs a stacked auto-encoder (SAE), and use a Deep Belief Network (DBN) to optimize the polarimetric feature sets. A key discovery of this paper is that even given an insufficient amount of data samples, deep learning allows achieving better performance than traditional algorithms by initializing its weights in a region near its local minimum with Stacked auto-encoders. Moreover, deep learning algorithms have very strong capabilities for exploring the complex correlation between features and obtain very promising fitting results on complicated data. [8] reports that oil slick classification achieved by deep networks outperformed both support vector machine (SVM) and traditional Artificial Neural Networks (ANN) with similar parameters. These studies consequently confirm the general trend towards successful deep learning, which is mainly related to the availability of large annotated databases to efficiently learn features from the data, as explained in [2].

ii. Fully Automatic Approaches

Another category of approaches proposes to solve the task with a deep neural architecture trained end to end. However, because of the lack of training data, transfer learning is generally considered and therefore reuses neural structures trained for another task and data that share some common features with the target context. Recent works on oil slick detection and segmentation outperform previous ones [25] [9]. This is in line with initial observations in the multimedia field which show the interest of transfer based approaches when not much data is available in the target domain and to reduce computation costs [5, 28].

Furthermore, [37] shows that initializing a network with transferred features from almost any number of layers can produce a boost to a generalization that persists even after fine-tuning to the target dataset. In this work, we consider transfer and domain adaptation of the Mask-RCNN, instance detection, and segmentation model [15] described in the next section. The full benefit will be derived from the huge annotated SAR sea surface datasets captured by different SAR sensors in different geographic areas. The diversity of the available data is large and rich enough to study the impact of the data behaviors and the learning strategies on detection and segmentation performance as well as the generalizability of the trained model.

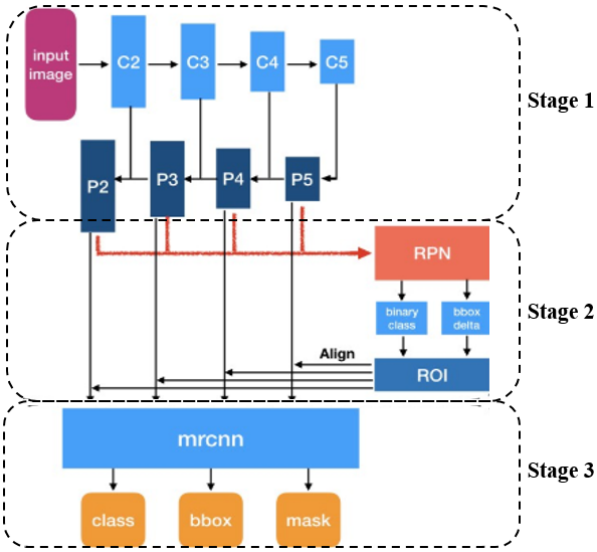


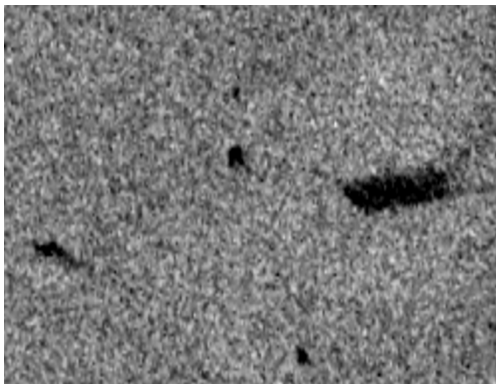
Figure 2: *Mask-RCNN architecture schema, from [39]*

III. Oil slick segmentation : Mask-RCNN

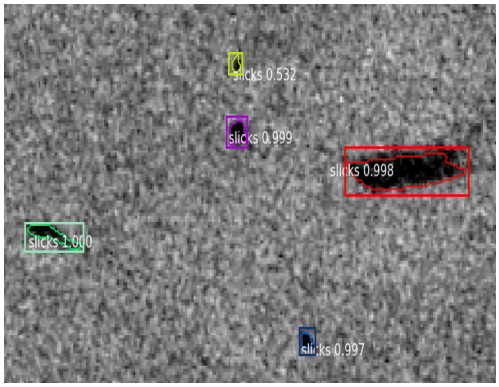
MaskRCNN is an instance segmentation method that combines object detection, classification. It goes one step further than classical semantic segmentation methods such as Fc-DenseNet [16] which are limited to pixel-level classification without differentiating between object instances. Mask-RCNN relies on a multi-stage convolutional network and inherits from the R-CNN series: R-CNN, Fast R-CNN, and Faster R-CNN [27]. It also competes with single-stage networks as YOLO [26] for the detection of large objects. But when detecting small objects, single-stage frameworks are generally much less efficient than two-stage frameworks [20]. This model is illustrated in Fig.2, can be summarized in three stages. The first step is a feature extraction step performed by a basic neural network that is usually pre-trained for an image classification task. Providing rich and transferable features that feed the higher stages. A second step is the Region Proposal Network (RPN) that generates areas of interest at different image scales. The final step is a multi-headed neural network that predicts the class of the object, refines the bounding box, and generates the associated object mask, as shown in Fig.3. This architecture is of real interest since it manages the extreme foreground-vs-background class imbalance through the selection process introduced by the RPN model. Therefore, it is well suited to the task of detecting and segmenting rare oil slicks. In the following section, we will briefly describe the main components of Mask-RCNN architecture.

i. Backbone Architecture and Region Proposal Network

The backbone is a Feature Pyramid Network (FPN). It extracts rich semantic features at all scale levels, combining low-resolution semantically strong features with high-resolution semantically weak features via a top-down pathway and lateral connections. The multi-scale description capability is especially important as the slicks have the property of having different sizes and shapes. This fact typically raises a huge challenge for deep learning that is only slightly invariant to scale. For the multimedia object detection challenge, Mask-RCNN relies on a backbone that is pre-trained on a classification problem. It is not possible, however, to provide such pre-training on SAR data at this time.



(a) Input Envisat image



(b) Resulting bounding box,associated masks and classes.

Figure 3: *Example of application of the Mask-RCNN approach on an Envisat SAR image(2002).*

ii. Head Architecture

We then propose to consider transfer learning relying on a multimedia pre-trained backbone, regardless of the difference between RGB and SAR images. The first layers of a deep neural network extract generic features that can be adapted to a variety of image analysis problems [37]. However, due to the discrepancy between multimedia RGB images and SAR images, a domain adaptation is required to fine-tune the backbone to our problem at a low learning rate. We rely on ResNet (50 or 101) backbones that improved the original version of Mask-RCNN [32]. Besides, ResNet is also recognized as a good choice in a variety of transfer case studies [32]. This raises the question: can we rely on a pre-trained network to transfer learned features from one domain to another and perform domain adaptation to obtain relevant features and perform accurate detection?

IV. SAR data sets description and performance assessment

In the present work, we consider Synthetic Aperture Radar (SAR) data from both Envisat and Sentinel-1 sensors. Some of the main SAR characteristics are the following: it relies on electromagnetic scattering, it provides high-resolution and large-sized data and it is altered by speckle noise [35]. These characteristics make it challenging to interpret SAR images and detect the oil targets, consuming a lot of manual work [4]. However, the huge amount of annotated data collected along time can be used to train an automatic detector in a supervised way.

i. Data Acquisition

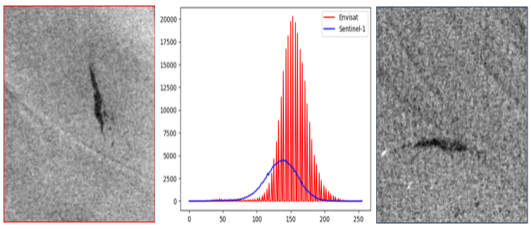
The methodologies and the results presented in the following are based on a database of about two thousand SAR images collected between the years 2002 and 2019 from the European Space Agency (ESA) missions: Envisat and Sentinel-1. These images are acquired in several areas along the acquisition period, mainly near Africa, which represents various geographical and meteorological contexts. Table 3 summarizes the SAR data used, their spatial resolution, and the acquisition period.

Table 3: *SAR Sensors.*

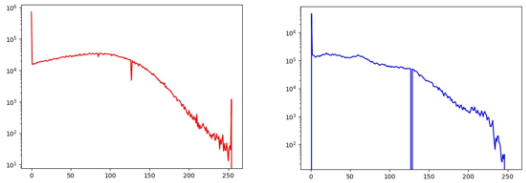
Sensor	Pixel Spacing(m)	Acquisition Period
Envisat	75	2002-2019
Sentinel-1	10	2002-2019

Fig.4 illustrates the pixel value distributions difference between slick and the other areas, for both Envisat and Sentinel-1 data. These curves are computed on interpreted SAR images from Envisat and Sentinel-1 images from

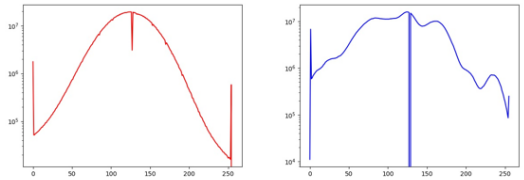
4 study areas. Fig.4(a) shows the histograms of two SAR images containing slicks (Envisat (left) and Sentinel (right)). We can observe that the shape of Envisat's pixel histogram (in red) is different from that of Sentinel-1 (blue). Besides, as shown in Fig.4 (b,c) when comparing the histograms of slick and not slick pixels for both sensors, one observes that the distributions are different. One could then expect that a model dedicated to a given sensor is not directly applied to the images produced by the other sensor.



(a) Histogram of slick image.



(b) Slick pixels histogram.



(c) Slick free pixels histogram.

Figure 4: *Difference between the histogram of Envisat(left)/Sentinel-1(right).*

ii. Human Experts Annotations

The remote sensing specialists at Total company have provided manual detection of the oil slicks, where each slick is labeled. These operators are experts, trained on the task and can make the difference between natural oil slicks, spills and lookalikes based on the visual identification of dark areas. The slick assessment is based on SAR images and external information used as a support during the analysis such as wind speed and nearby oil rigs and ships. Fig.5 shows an association of SAR data with an example of slick annotation of image interpreters. The red annotations present seeps and the blue one's present spills. Note that lookalikes can be seen but are not annotated by experts.

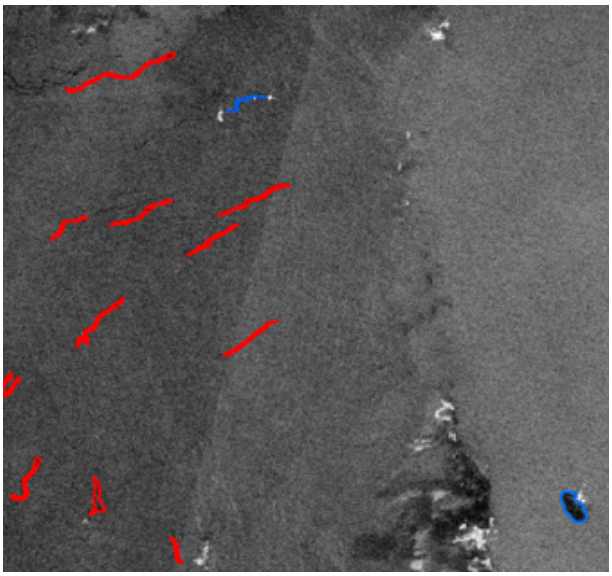


Figure 5: *Example of manual annotation of image interpreters on an Envisat SAR 2002 image, oil seeps (red) and oil spill (blue)*

iii. Datasets Pre-Processing

From raw SAR images to SAR images provided to human experts or machine learning algorithms, processing flow is established and illustrated in figure 6. The related sub processes are summarized as follows:

Pre-processing is a necessary step of pre-processing applied to the raw N1 images. Pre-processing consists of transformations of low-level SAR data to improve the qualitative and quantitative interpretation of image components. Integrated into a standardized pipeline, it includes geometric, radiometric and atmospheric corrections, as well as intensity level correction. The geo-referencing

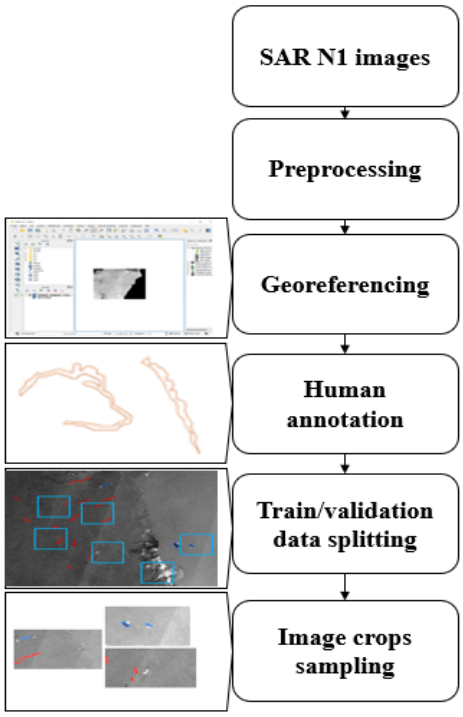


Figure 6: *Flowchart for SAR data pre-processing*

step comes afterward, it is an important step that realizes the correspondence with the requirements of GIS (geo-referenced information systems) [12].

Ground Truth (GT) Integration: aims to associate each SAR image with the corresponding human expert slick annotations to serve both the training and validation steps. A class label is associated with each pixel in the SAR image and all instances of the slick are differentiated.

Training and Validation Datasets Splitting: involves precise data selection. Two strategies are tested, the first is to select N geographic areas (called studies) for the training and M others for the validation dataset. The second consists in mixing the geographical areas in the two sets of training and validation (but avoiding the corresponding image redundancies). The objective is to understand the impact of geographic variability on the generalizability of the model. Note that ground and coastal areas are removed thus avoiding ground patterns learning.

Image Crops Sampling: the last step is to create image crops based on the large SAR images. Being aware that a neural network has a limited field of view and requires a large amount of memory, the size of the crop must be larger than the receiving field while still allowing the model to fit into a given processing device, typically a GPU with 16 Gb of local memory. Besides, the selected crops must present the diversity of data to the network while maintaining a certain balance between classes. Then, crop selection is based on random region sampling in large images. These crops are then checked by filters. All the crops containing a given ratio of slick pixels are preserved while only a few free slick crops are selected if they contain probable lookalikes. Since these areas are not annotated but have a higher variance of pixel values than the other free areas, their selection is based on a minimum SAR pixel variance filter. Since these areas are not annotated but have a higher variance of pixel values than the other free areas, their selection is based on a minimum SAR pixel variance. A random selection is then applied to limit their quantity compared to slicks crop in the dataset. Data augmentation is performed in order to artificially increase the variability of the data. It consists of random horizontal and vertical flips [34] and noise addition. This noise is a null Gaussian mean, it effectively distorts the characteristics of high-frequency elements. Learning the latter can generate an over-fitting [23].

iv. Performance Metrics

We consider a task of object instance detection and segmentation. Given our slick instance-level annotated datasets, we rely on the classical metrics in the domain. The notation of the metrics used for the evaluation is as follows : False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN). The following metrics are calculated :

- Intersection-over-Union (IoU): measures the compliance between the masks positioning and size.
  $IoU(Y, \hat{Y}) = \frac{TP}{TP+FN+FP} = \frac{Area\ of\ Overlap}{Area\ of\ Union}$
  , where Y is the prediction and $\hat{Y}$ is the Ground truth (GT). The IoU can also serve as a good indicator for segmentation, but it cannot tell us how good are the obtained detection results. We consider that the slick is detected even with a low IoU score (partial detection).
- The pixel confusion matrix.

V. Experimental Design

Our experiment plan has two main objectives, where the first is to study the impact of the training/validation dataset selection strategy. And the second is to study the network architecture, hyperparameters and learning strategies.

i. Selection strategy of training/validation datasets

The last two steps of the data processing shown in Fig. 6 are set to study the impact of the following parameters :

- Selection of SAR data sensors : Envisat and Sentinel-1 have different sensors behaviors such as the resolution and pixel value distributions.
- Selection of data acquisition areas : each area represents different geographical and meteorological contexts. The characteristics of the slick are directly affected by meteorological conditions [13].
- Control on the introduction of lookalike phenomena : either avoid them or introduce a given amount in the training data set.

ii. Network and Architecture Parameters

We consider the following parameters:

- Network Architecture : we experiment with the Mask-RCNN model different backbone architectures (feature map extractors): ResNet with either 50 or 101 layers.
- Loss Function : an extreme imbalance between foreground (slick) and background (sea) classes during training is observed as shown in Table 1. To down-weight the easy examples and to focus the training on the hard negatives, we use the focal loss [18] instead of the Cross-Entropy(CE) for the mask loss computation. A modulating factor is added to the CE loss where integer λ is defined as 2 in the experiment.

$$L(p_t) = -\alpha_t(1-p_t)^{\lambda} log(p_t) \qquad (1)$$

  where p_t is the predicted probability of the class. The role of the parameters α_t and λ is to down-weight the easy examples and thus focus the optimization on the hard negatives [18]. When λ = 0, the focal loss is equivalent to the EC. Increasing λ also increases the effect of the modulation factor. In our case, we've set it at 2.

VI. Results and Discussion

Fig.7 illustrates the results of the detection on various SAR images including the presence of various lookalike phenomena. The model considered: relies on the Resnet-101 backbone and trained on data from mixed study areas (training/validation), it is trained on Envisat data with the addition of lookalike samples. In these images, the SAR information is displayed in red, the expert annotation (GT) in blue and the prediction in green. The GT overlay associated with good prediction is shown in light blue. Visual inspection shows that only a few lookalikes are detected as slicks (the green color highlights pixels predicted as slick but are not slick i.e. False alarms). Most of the slicks are partially detected, highlighting issues related to slick boundaries; either the detected slick is outside GT or is not entirely detected. Almost all the slicks are detected with a classification

score higher than 0.9. Tab.4 shows the confusion matrix of the image in the upper right corner of Fig.7. It shows that 1063 pixels are detected as a slick and they are slick pixels(light blue color), 281 pixels are detected as slick and they are sea pixels, 1187 sea pixels are detected as slick (green color). The subsequent values are mainly the result of erroneous predictions on the slick boundaries.

Table 4: *The confusion matrix in pixels numbers.*

		Prediction	
		Slick	Sea
Ground Truth	Slick	1063	1187
	Sea	281	259613

An analysis of performance is detailed in the sequel. We focus our attention only on the most relevant results, the quantitative measures of our analysis are reported in Table 5. Experiments have been carried out on both Envisat and Sentinel-1 satellite images. A first observation is that absolute performance levels are different between sensors, due to the change in their data distributions and their different characteristics mentioned in Table 3. however, to preserve the readability of the paper, we will limit the presentation of the results to a specific sensor.

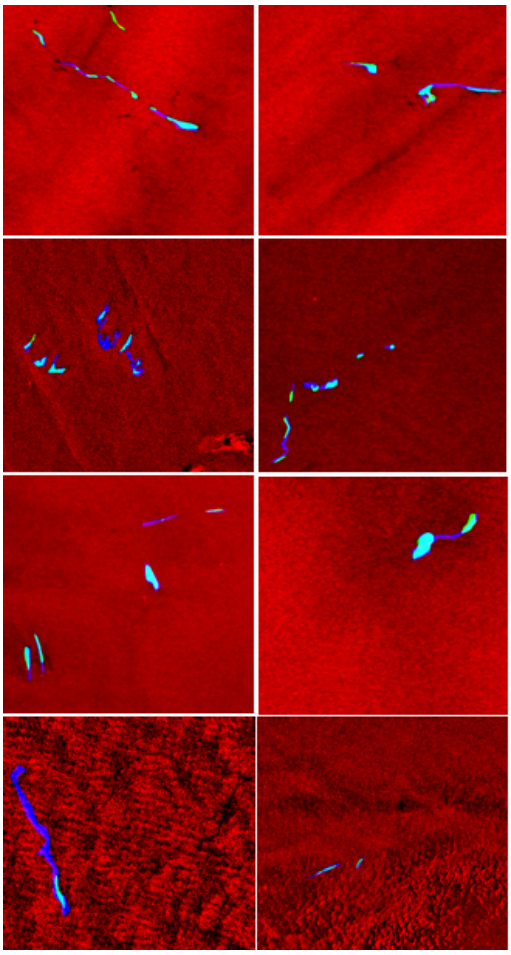


Figure 7: *Example of a prediction result on Envisat SAR images from a set of tests, good predictions (light blue), false alarms (green), non-detections (dark blue).*

- In general, and as shown in Fig.7, there is an annotation uncertainty about the slick boundaries at the manual annotation step. Therefore results cannot be expected to reach maximum performance measure values. A maximum mean IoU (mIoU) value of 0.65 is obtained in our experiments. Finding the border of the objects was indeed the most difficult problem.This manifests itself in missing detection and misadjusted edge.
- Introducing a given number of lookalikes in the training data improves the IoU slick by 42%, yielding less over-fitting, as shown in Table 5(Envisat A). The first phenomenon targeted by the network when over-fitting is lookalikes, which are much more important than the slick ones. Then, their explicit addition in the train dataset helps avoid them.
- The use of focal loss instead of CE for the calculation of mask loss shows a slight improvement, where a 19% mIoU improvement is reported in Table 5(Envisat B). Focal loss compensates for class imbalance and refines slicks boundaries.
- Table 5(Envisat C) indicates the impact of the slick position inside the crop. The mIoU improves by 23% in the mIoU when randomizing crops position. The sea IoU decreases

by 19 % in the case of generation of centered slicks with more false alarms, this may be due to a bias in the RPN module caused by the systematic centering of the slick.
- The results indicates that the generalization of the test set is considerably improved when training and validation take into account mixed images from different study areas as shown in Table 5(Envisat D). The results are better than in the case where training and validation are carried out on different studies from different areas.
- Using the deeper ResNet backbone (Resnet-101) architecture yields better mIoU performance as shown in Table 5(Envisat E). Conclusions are similar to [24].
- The ImageNet pre-trained weights yield a slightly better result than the Coco pre-trained weights. This may be due to the variety in the object sizes in the ImageNet dataset as shown in Table 5(Sentinel-1 F).
- Networks trained on Envisat data perform slightly better than those trained on Sentinel-1 data. This can be observed by comparing Envisat based measurements with those based on Sentinel-1 as shown in Table 5. However, the sensors areas and pre-processing are different and Envisat benefit from a longer experience such that a more dedicated study must be conducted before drawing more comparison.
- As illustrated in Fig.7, a large slick may be detected as different slick instances. Different hyper-parameters impact on this issue, mainly the mask shape size and the RPN anchor scales. Morphological post-processing can be carried out to optimize large target detection.

Table 5: *Multi-criteria performance measures on the test dataset for different configurations*

Name	IoU Sea	IoU Slick	mIoU
Envisat A			
With lookalikes	0.84	0.5	0.52
Without lookalikes	0.99	0.35	0.65
Envisat B			
Focal loss	0.95	0.49	0.62
CE loss	0.83	0.49	0.52
Envisat C			
Centered slick	0.83	0.49	0.52
Random slick	0.99	0.32	0.64
Envisat D			
Mixed areas	0.88	0.49	0.56
Not mixed areas	0.73	0.49	0.45
Envisat E			
ResNet101	0.99	0.35	0.65
ResNet50	0.99	0.02	0.51
Sentinel-1 F			
COCO weights	0.97	0.25	0.56
ImageNet weights	0.97	0.27	0.56

Experimental comparison results are obtained during training with specific hyper-parameters. Envisat A mixes study areas or restricts training to specific areas, Envisat B shows the impact of the ResNet backbone complexity, Envisat C indicates the impact of introducing lookalikes in the training data, Envisat D compares the effect of changing the mask loss, Envisat E presents the effect of slick position within the images and Sentinel-1 E reports the effect of the pre-trained weights.

VII. Conclusion and Future works

Introducing deep learning in the field of HSE(Health Security Environment) field is a research trend. In this paper, a specific effort has been placed to study the deep SAR data instance segmentation method. We demonstrate that the Mask-RCNN instance segmentation approach, although primarily designed with object detection, object localization and segmentation of natural image instances, can be used to produce a promising ability for automatic detection of the offshore oil slick, under different weather conditions and in a variety of locations.

Our future work will focus on three main directions: (1) Improving the dataset and its efficiency by adding more useful information such as meteorological information and oil platforms and pipeline position. (2) A full analysis of model optimization and limitation

of over-fitting and edge problems. (3) An exploration of the performance not only for detecting but also characterizing offshore oil slick(seep, spill).

VIII. Acknowledgments

This work has been carried out thanks to the computing facilities and annotated data provided Total company.

References

[1] Shivang Agarwal, Jean Ogier Du Terrail, and Frédéric Jurie. Recent advances in object detection in the age of deep convolutional neural networks. *arXiv preprint arXiv:1809.03193*, 2018.

[2] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.

[3] Werner Alpers, Benjamin Holt, and Kan Zeng. Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sensing of Environment*, 201:133–147, 2017.

[4] Sébastien Angelliaume, Pascale C Dubois-Fernandez, Cathleen E Jones, Benjamin Holt, Brent Minchew, Emna Amri, and Véronique Miegebielle. Sar imagery for detecting sea surface slicks: Performance assessment of polarization-dependent parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[5] Yoshua Bengio, Frédéric Bastien, Arnaud Bergeron, Nicolas Boulanger-Lewandowski, Thomas Breuel, Youssouf Chherawala, Moustapha Cisse, Myriam Côté, Dumitru Erhan, Jeremy Eustache, et al. Deep learners benefit more from out-of-distribution examples. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 164–172, 2011.

[6] Camilla Brekke and A.H. Solberg. Classifiers and confidence estimation for oil spill detection in envisat asar images. *IEEE Geoscience and Remote Sensing Letters*, 5(1):65–69, 2008.

[7] Camilla Brekke and Anne HS Solberg. Oil spill detection by satellite remote sensing. *Remote sensing of environment*, 95(1):1–13, 2005.

[8] Guandong Chen, Yu Li, Guangmin Sun, and Yuanzhi Zhang. Application of deep networks to oil spill detection using polarimetric synthetic aperture radar images. *Applied Sciences*, 7(10):968, 2017.

[9] Cheng Siong Chin, JianTing Si, AS Clare, and Maode Ma. Intelligent image recognition system for marine fouling using softmax transfer learning and deep convolutional neural networks. *Complexity*, 2017, 2017.

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[11] Antonio De Maio, Giuseppe Ricci, and Manlio Tesauro. On cfar detection of oil slicks on the ocean surface by a multifrequency and/or multipolarization sar. In *Proceedings of the 2001 IEEE Radar Conference (Cat. No. 01CH37200)*, pages 351–356. IEEE, 2001.

[12] Merv Fingas and Carl Brown. Review of oil spill remote sensing. *Marine pollution bulletin*, 83(1):9–23, 2014.

[13] Mervin F Fingas and Carl E Brown. Review of oil spill remote sensing. *Spill*

Science & Technology Bulletin, 4(4):199–208, 1997.

[14] Hao Guo, Danni Wu, and Jubai An. Discrimination of oil slicks and lookalikes in polarimetric sar images using cnn. *Sensors*, 17(8):1837, 2017.

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.

[16] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1175–1183. IEEE, 2017.

[17] Thomas FN Kanaa, E Tonye, G Mercier, V de P Onana, and JP Rudant. Multiscale segmentation of oil slick in sar images based on morphological pyramid. In *ENVISAT and ERS Symposium, Salsburg, Australie*, pages 6–10, 2004.

[18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[19] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.

[20] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, pages 1–58, 1809.

[21] Peng Liu, Ying Li, Bingxin Liu, Peng Chen, and Jin Xu. Semi-automatic oil spill detection on x-band marine radar images using texture analysis, machine learning, and adaptive thresholding. *Remote Sensing*, 11(7):756, 2019.

[22] David Mera, José M Cotos, José Varela-Pet, and Oscar Garcia-Pineda. Adaptive thresholding algorithm based on sar images and wind data to segment oil spills along the northwest coast of the iberian peninsula. *Marine pollution bulletin*, 64(10):2090–2096, 2012.

[23] Arvind Neelakantan, Luke Vilnis, Quoc V Le, Ilya Sutskever, Lukasz Kaiser, Karol Kurach, and James Martens. Adding gradient noise improves learning for very deep networks. *arXiv preprint arXiv:1511.06807*, 2015.

[24] Xuan Nie, Mengyang Duan, Haoxuan Ding, Bingliang Hu, and Edward K Wong. Attention mask r-cnn for ship detection and segmentation from remote sensing images. *IEEE Access*, 8:9325–9334, 2020.

[25] Teerapong Panboonyuen, Kulsawasd Jitkajornwanich, Siam Lawawirojwong, Panu Srestasathiern, and Peerapon Vateekul. Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning. *Remote Sensing*, 11(1):83, 2019.

[26] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[28] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and

Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[29] AH Schistad Solberg, Geir Storvik, Rune Solberg, and Espen Volden. Automatic detection of oil spills in ers sar images. *IEEE Transactions on geoscience and remote sensing*, 37(4):1916–1924, 1999.

[30] Anne H Schistad Solberg and Espen Volden. Incorporation of prior knowledge in automatic classification of oil spills in ers sar images. In *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing-A Scientific Vision for Sustainable Development*, volume 1, pages 157–159. IEEE, 1997.

[31] HIM Struckmeyer, AK Williams, R Cowley, JM Totterdell, G Lawrence, and Geoffrey William O'Brien. Evaluation of hydrocarbon seepage in the great australian bight. *The APPEA Journal*, 42(1):371–385, 2002.

[32] Sasha Targ, Diogo Almeida, and Kevin Lyman. Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*, 2016.

[33] K Topouzelis, V Karathanassi, P Pavlakis, and D Rokos. Detection and discrimination between oil spills and look-alike phenomena through neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(4):264–270, 2007.

[34] Jason Wang and Luis Perez. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 2017.

[35] Puyang Wang, He Zhang, and Vishal M Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017.

[36] SY Wu and AK Liu. Towards an automated ocean feature detection, extraction and classification scheme for sar imagery. *International Journal of Remote Sensing*, 24(5):935–951, 2003.

[37] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328, 2014.

[38] Victoria V Zatyagalova, Andrei Yu Ivanov, and Boris N Golubov. Application of envisat sar imagery for mapping and estimation of natural oil seeps in the south caspian sea. In *Proceedings of the Envisat symposium, Montreux, Switzerland (ESA SP-636*, 2007.

[39] Xiang Zhang. Simple Understanding of Mask RCNN. `https://medium.com/@alittlepain833/simple-understanding-of-mask-rcnn-134b5b330e95`, 2020. [Online; accessed 04-04-2020].