# Compositional Oil Spill Detection Based on Object Detector and Adapted Segment Anything Model from SAR Images

Wenhui Wu, Man Sing Wong, Xinyu Yu, Guoqiang Shi, Coco Yin Tung Kwok, and Kang Zou

Abstract-Semantic segmentation-based methods have attracted extensive attention in oil spill detection from SAR images. However, the existing approaches require a large number of finely annotated segmentation samples in the training stage. To alleviate this issue, we propose a composite oil spill detection framework, SAM-OIL, comprising an object detector (e.g., YOLOv8), an Adapted Segment Anything Model (SAM), and an Ordered Mask Fusion (OMF) module. SAM-OIL is the first application of the powerful SAM in oil spill detection. Specifically, the SAM-OIL strategy uses YOLOv8 to obtain the categories and bounding boxes of oil spill-related objects, then inputs bounding boxes into the Adapted SAM to retrieve category-agnostic masks, and finally adopts the OMF module to fuse the masks and categories. The Adapted SAM, combining a frozen SAM with a learnable Adapter module, can enhance SAM's ability to segment ambiguous objects. The OMF module, a parameterfree method, can effectively resolve pixel category conflicts within SAM. Experimental results demonstrate that SAM-OIL surpasses existing semantic segmentation-based oil spill detection methods, achieving mIoU of 69.52%. The results also indicated that both OMF and Adapter modules can effectively improve the accuracy in SAM-OIL.

Index Terms—Oil Spill Detection, Object Detection, Segment Anything Model, Adapter.

#### I. Introduction

YNTHETIC Aperture Radar (SAR) is one of the most efficient remote sensing tools for detecting oil spills [1], [2]. One challenge in oil spill detection is to distinguish between oil spills and look-alikes, as both appear as dark spots in SAR imagery [1]–[3]. Dong *et al.* [4] analyzed 563,705 Sentinel-1

Manuscript received 15 January 2024; revised 3 March 2024; accepted 14 March 2024. Date of publication XXX; date of current version XXX. This work is funded by Environment and Conservation Fund. Any opinions, findings, conclusions or recommendations expressed in this material/event do not necessarily reflect the views of the Government of the Hong Kong Special Administrative Region and the Environment and Conservation Fund and the 'Woo Wheelock Green Fund' (for jointly funded project). M.S. Wong also thanks the support from the General Research Fund (project ID: 15609421 and 15603920), the Research Grants Council of Hong Kong. (Corresponding author: Man Sing Wong)

M.S. Wong is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, and also with the Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: Ls.charles@polyu.edu.hk).

W. Wu, X. Yu, G. Shi, C.Y.T. Kwok, and K. Zou are with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: wenhui.wu@polyu.edu.hk; 19109597r@connect.polyu.hk; guoqiang.shi@polyu.edu.hk; yt-coco.kwok@connect.polyu.hk; welly.zou@polyu.edu.hk).

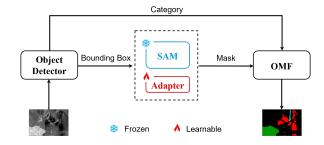


Fig. 1. Architecture of the proposed method. The Adapted SAM is composed of a frozen SAM and a learnable Adapter module. SAM-OIL consists of an object detector, an Adapted SAM, and an OMF module.

images and found that about 90% of oil slicks were located within 160 kilometers of shorelines, 21 among them are high-density slick belts which show similar to shipping routes. Therefore, shorelines and ships serve as crucial contextual information for oil spill detection and can assist in determining the source of oil spills [3], [4]. Krestenitis *et al.* [5] developed a multi-class semantic segmentation dataset, which includes oil spill, look-alike, land, ship, and sea surface. In this letter, we cite this dataset as the Multimodal Data Fusion and Analytics (M4D) dataset, following the terminology of the proposing team's institution.

The traditional approach to detect oil spills on SAR images consists of three steps: dark spot segmentation, feature extraction, and distinguishing oil spills from look-alikes [1], [2]. Firstly, dark spots on the SAR image are extracted through traditional segmentation methods. Then, various features such as geometric, contextual, texture, and physical features are extracted from these dark areas. Finally, a traditional classifier such as Support Vector Machine or Random Forest is used to distinguish oil spills from look-alikes [6]. However, these methods have certain limitations, such as issues of undersegmentation or over-segmentation during the segmentation process, limited expressive capability of manually designed features, and the relatively weak discriminative ability of traditional classifiers in identifying oil spills. Therefore, there is still room for improvement in the accuracy of oil spill detection.

Deep learning approaches input with raw images and learn hierarchical feature representations from shallow visual features to deep semantic features from a large set of data. This process reduces the dependence on manually processing, demonstrating a strong capability to detect and capture complex patterns [2]. Therefore, deep learning-based methods have gained increasing attention in oil spill detection recently [5], [7]–[9]. Previous studies either directly used classical semantic segmentation models to detect oil spills, or developed tailor-made adaptive semantic segmentation models based on the characteristics of oil spills for the detection [5], [7]. In addition, the integration of polarimetric features in these models further enhances detection accuracy [8], [9]. Although these methods provided reasonably good detection performance, they required accurately annotated segmentation samples for training, which is time-consuming and resource-intensive for building the dataset.

To reduce the dependency on extensively annotated samples for semantic segmentation, we introduce SAM-OIL, a composite oil spill detection framework that includes an object detector, an Adapted Segment Anything Model (SAM), and an Ordered Mask Fusion (OMF) module. SAM represents the latest advancements in natural image segmentation, which offers impressive zero-shot segmentation performance by using input prompts such as points, bounding boxes, masks, and text [10]. Among the types of prompts, those based on bounding boxes have led to the best segmentation results [11]. Hence, we integrate an object detector (e.g., YOLOv8 [12]) with SAM for oil spill detection. On the other hand, the outputs of SAM are category-agnostic binary masks, which may cause pixel category conflicts when masks are merged. To address this issue, we proposed the Ordered Mask Fusion (OMF) module, which merges masks according to a predefined category order. Furthermore, directly applying the SAM trained on natural images to remote sensing satellite imagery does not yield satisfactory results [11], especially for SAR images with blurred object boundaries. To bridge this gap, SAM needs to be adapted specifically for oil spill detection. HQ-SAM [13] is the minimal adaptation of SAM, improving its segmentation precision by adding just one Adapter module, which accounts for no more than 0.5% of SAM's parameters. Due to the simplicity and training efficiency of HQ-SAM, we integrate the Adapter module from HQ-SAM into SAM-OIL. The main contributions of this study are summarized as follows.

- We propose a compositional oil spill detection framework called SAM-OIL, which combines an advanced object detector (e.g., YOLOv8) with the SAM. To our best knowledge, this is the first time the SAM has been applied to oil spill detection. Furthermore, SAM-OIL framework can reduce the need for finely annotated samples.
- Ordered Mask Fusion (OMF) module is proposed, which is a parameter-free method, to effectively resolve pixel category conflicts in SAM.
- We introduce an Adapter module from HQ-SAM into the SAM-OIL, which utilizes masks from M4D dataset to train the Adapter, significantly enhancing SAM's segmentation capabilities for objects with fuzzy boundaries.
- Experimental results demonstrate that SAM-OIL achieves mIoU of 69.52%, surpassing existing oil spill detection methods, and both OMF and Adapter can successfully improve the accuracy of SAM-OIL.

The rest of this letter is organized as follows. Section II describes the methods proposed. Section III performs a series of quantitative and qualitative analyses via experiments. Section IV concludes this letter.

#### II. METHODOLOGY

The architecture of the proposed method is given in Fig. 1. SAM-OIL consists of an object detector (i.e., YOLOv8), an Adapted SAM, and an OMF module. We refer to the combination of YOLOv8 and frozen SAM as YOLOv8-SAM, which is considered as the baseline of our method.

## A. YOLOv8-SAM

SAM consists of three components: an image encoder  $\Phi_{enc-i}$ , a prompt encoder  $\Phi_{enc-p}$ , and a lightweight mask decoder  $\Phi_{dec-m}$ . SAM takes an image I and a set of prompts P as inputs, which first uses  $\Phi_{enc-i}$  to obtain image features  $F_I$ , and encodes the prompts P into sparse prompt tokens  $T_P$  through  $\Phi_{enc-p}$ . Then,  $T_P$  and some learnable output tokens  $T_O$  are concatenated, and finally input into  $\Phi_{dec-m}$  along with  $F_I$  for attention-based feature interaction, generating categoryagnostic binary masks M. This process can be described as:

$$\begin{cases}
F_I &= \boldsymbol{\Phi}_{enc-i}(I) \\
T_P &= \boldsymbol{\Phi}_{enc-p}(P) \\
M &= \boldsymbol{\Phi}_{dec-m}(F_I, Concat(T_O, T_P))
\end{cases} \tag{1}$$

YOLOv8-SAM is composed of an object detector and a frozen SAM. The procedure is as follows: 1) The object detector  $\Phi_{det}$  obtains the object categories  $C_{bbox}$ , bounding boxes  $P_{bbox}$ , and classification score  $S_{bbox}$ .  $P_{bbox}$  is filtered based on the defined threshold  $S_{threshold}$  for  $S_{bbox}$ . 2) The image I and a set of prompts  $P_{bbox}$  are input into SAM, to obtain prompt tokens  $T_{bbox}$  and masks  $M_{bbox}$ . 3) The  $M_{bbox}$  and the corresponding  $C_{bbox}$ , are randomly merged to obtain the final masks  $M_{semantic}$ . Considering the effectiveness of the YOLO series models, we choose the latest YOLOv8 as the object detector. The complete process is described as follows:

$$\begin{cases}
C_{bbox}, P_{bbox} &= \begin{cases} \mathbf{\Phi}_{det}(I) \\ S_{bbox} > S_{threshold} \end{cases} \\
F_{I} &= \mathbf{\Phi}_{enc-i}(I) \\
T_{bbox} &= \mathbf{\Phi}_{enc-p}(P_{bbox}) \\
M_{bbox} &= \mathbf{\Phi}_{dec-m}(F_{I}, Concat(T_{O}, T_{bbox})) \\
M_{semantic} &= Random(C_{bbox}, M_{bbox})
\end{cases}$$
(2)

# Algorithm 1 Pseudocode of OMF in a Numpy-like style.

```
# Detector outputs categories, SAM outputs masks.
classes, bboxes, scores = Detector(img)
masks = SAM(img, bboxes)

# Pre-defined category fusion order.
order = ["ship", "land", "oil_spill", "look-alike"]

# Sort masks according to the 'order'.
zipped = zip(classes, masks)
sorted_masks = sorted(zipped, order)

# Fuse the masks.
result = zeros((img_h, img_w, 1))
for mask in sorted_masks:
    result[result==0] = mask[result==0]
```

## B. Ordered Mask Fusion module

The Ordered Mask Fusion (OMF) module is proposed to address pixel category conflicts encountered in SAM's mask fusion by using a specific order for masks, as detailed in Algorithm 1. The order determines the categorization of ambiguous pixels. For instance, if the order is "ship, land, oil spill, look-alike", and a pixel is identified as both "ship" and "land", it will be classified as "ship".

## C. Adapted SAM

We adopt HQ-SAM [13] as the Adapted SAM because of its simplicity and training efficiency. The HQ-SAM is composed of a frozen SAM and a learnable Adapter module. During the training process, all the pre-trained SAM parameters are frozen, while only the Adapter module is updated. Adapter module increases  $\leq 0.5\%$  of SAM's parameters, thereby ensuring the module's efficiency in training [13]. The HO-SAM was initially developed to enhance SAM's segmentation accuracy for intricate structures in natural images, while the HQ-SAM used in this study aims to improve SAM's ability to segment blurry boundaries in SAR imagery. The Adapter module is trained for specific tasks like oil spill detection. Thus, using SAM-OIL with this module may change or decrease accuracy in other tasks, but without affecting SAM's generalization. Additionally, the Adapter module can be retrained for different applications.

HQ-SAM was constructed by introducing the fusion of HQ-Output Token and deep-shallow feature into SAM. Firstly, a learnable HQ-Output Token  $T_{HQ}$  is designed, which is input into SAM's mask decoder  $\Phi_{dec-m}$  together with the original prompts  $T_P$  and output token  $T_O$ . Unlike the original output token  $T_O$ , the HQ-Output Token  $T_{HQ}$  and its related Multilayer Perceptron (MLP) layer are trained to predict highquality segmentation masks. Secondly, by fusing the early feature  $F_{early}$  and late feature  $F_I$  of its image encoder, it uses both global semantic context and local fine-grained features. We refer to the combination of the HQ-Output Token, threelayer MLP, and a small feature fusion block as the Adapter module of HQ-SAM. The combination of the Adapter module and the original SAM decoder  $\Phi_{dec-m}$  forms a new decoder  $\Phi'_{dec-m}$ . In summary, the SAM-OIL is composed of an object detector, HQ-SAM, and an OMF module, of which the whole process can be described as:

$$\begin{cases} C_{bbox}, P_{bbox} &= \begin{cases} \boldsymbol{\varPhi}_{det}(I) \\ S_{bbox} > S_{threshold} \end{cases} \\ F_{early}, F_{I} &= \boldsymbol{\varPhi}_{enc-i}(I) \\ T_{bbox} &= \boldsymbol{\varPhi}_{enc-p}(P_{bbox}) \\ T &= Concat(T_{HQ}, T_{O}, T_{bbox}) \\ M_{bbox} &= \boldsymbol{\varPhi}'_{dec-m}(F_{early}, F_{I}, T) \\ M_{semantic} &= \boldsymbol{\varPhi}_{OMF}(C_{bbox}, M_{bbox}) \end{cases}$$
(3)

# III. EXPERIMENTS

# A. Experimental Setups

1) Dataset: The M4D dataset [5] is used in this study, which is composed of five categories: oil spill, look-alike, land, ship,

and sea surface. It includes a total of 1,112 SAR images, each with a dimension of  $650\times1250$  pixels. The images were divided into training and testing sets, with 1,002 images (90%) for training and 110 images (10%) for testing.

3

- 2) Evaluation Metrics: We use Intersection over Union (IoU), mean Intersection over Union (mIoU) and mF1 (mean F1-score) for evaluation, commonly applied in semantic segmentation. IoU measures the accuracy of individual classes, while mIoU and mF1 assess overall accuracy.
- 3) Configuration Details: All the experiments were conducted on two NVIDIA TITAN RTX GPUs (48G memory), with the operating system of Ubuntu 18.04.6 LTS. The object detector used was YOLOv8-X in MMYOLO [12], configured for 1,000 training epochs (note, early stopping strategy was employed) and batch size of 6, with all other parameters kept at MMYOLO's default settings. We also used a pre-trained model on the COCO dataset to initialize YOLOv8-X parameters [12]. To maintain consistency with HQ-SAM's training scheme, the M4D dataset was refined by creating a binary mask for each object, without including any additional class information. These binary masks were then used to train the HQ-SAM Adapter module. The training involved 120 epochs, using the original image size (650×1250) and bounding box prompts. The ViT-H [10] model was employed as the image encoder for the SAM. The remaining parameters followed the HQ-SAM's default configuration.
- 4) Comparative experiment setup details: Inspired by [5], we applied several classic semantic segmentation methods implemented in MMSegmentation [14] to the M4D dataset, including U-Net, PSPNet, UPerNet, DeepLabV3+, and OCRNet, with each model's backbone detailed in Table I. All comparative methods were conducted on the same hardware configuration as SAM-OIL, with training epochs set at 300, batch size adjusted according to model size, and all other training hyperparameters following MMSegmentation's default settings.

## B. Comparative Experiments

Table I compares the experimental results of our methods with other methods. SAM-OIL achieved 69.52% mIoU, higher than DeepLabV3+ (67.59%), by 1.93%. It also achieved the best results in the "Oil Spill" and "Ship" categories. Notably, in the "Ship" category, SAM-OIL reached IoU of 52.55%, higher than DeepLabV3+ (36.99%), by 15.56%. This indicates SAM-OIL, incorporating an object detector, has significant advantages in detecting small objects. Moreover, SAM-OIL outperforms our baseline model, i.e., YOLOv8-SAM, across all categories, demonstrating the effectiveness of the OMF and Adapter modules used (see the Ablation Study section for detailed analysis). However, in the "look-alike" category, SAM-OIL's IoU was 55.60%, lagging behind the best model, OCRNet (63.53%), by 7.93%. It still requires improvement.

Table I lists each model's parameters, showing that SAM-OIL has more than the others, which means it uses more computational resources and is slower in training and inference. Fig.2 shows that YOLOv8-SAM achieves the fastest training speed per epoch since it only trains YOLOv8, while SAM-OIL trains both YOLOv8 and the Adapter, leading to longer

TABLE I

QUANTITATIVE PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND THE SEMANTIC SEGMENTATION METHODS ON THE M4D

DATASET.

Methods	Backbone	Params.(M)	Sea Surface	Oil Spill	Look-alike	Ship	Land	mIoU	mF1
U-Net PSPNet	/ ResNet50	<b>28.99</b> 46.60	92.21 96.80	43.92 44.37	31.04 59.77	21.07 27.61	90.74 92.78	55.79 64.27	66.86 74.84
UPerNet	ResNet50	64.04	96.51	44.37	58.17	31.38	93.89	65.72	76.37
DeepLabV3+ OCRNet	ResNet50 ResNet50	41.22 70.35	97.01 <b>97.03</b>	49.09 50.07	61.79 <b>63.53</b>	36.99 32.65	93.06 <b>94.11</b>	67.59 67.48	78.22 77.82
YOLOv8-SAM	YOLOv8-X+ViT-H	68.16+2446	94.34	41.84	48.15	52.48	87.65	64.89	76.67
SAM-OIL	YOLOv8-X+ViT-H	68.16+2452.10	96.05	51.60	55.60	52.55	91.81	69.52	80.43

<sup>\*</sup> All results are described using percentages (%). The accuracy of each category is evaluated using IoU. The best score is marked in bold.

 $\label{thm:table II} \textbf{The influence of object detector on the accuracy}.$ 

 THE INFLUENCE OF OMF AND A	DAPTER O	N THE A	ACCURACY.
Methods		MF	Adanter

Methods	mIoU		
YOLOv8-SAM	64.89		
YOLOv8-SAM (gt_box)	77.09		
SAM-OIL	69.52		
SAM-OIL (gt_box)	80.35		

Methods	OMF	Adapter	mIoU
YOLOv8-SAM			64.89
YOLOv8-SAM + OMF	$\checkmark$		65.82
YOLOv8-SAM + Adapter		✓	67.97
SAM-OIL (YOLOv8-SAM + OMF + Adapter)	$\checkmark$	$\checkmark$	69.52

TABLE III

 $\label{thm:table_iv} \textbf{TABLE IV}$  The influence of fusion orders of object masks in OMF on the accuracy.

Orders	Sea Surface	Oil Spill	Look-alike	Ship	Land	mIoU
Random	96.05	48.34	54.87	50.10	90.47	67.97
Look-alike, Oil Spill, Ship, Land	96.05	48.31	54.84	41.78	90.41	66.28
Look-alike, Oil Spill, Land, Ship	96.05	48.31	54.84	41.78	90.41	66.28
Look-alike, Land, Oil Spill, Ship	96.05	48.31	54.84	41.78	90.41	66.28
Ship, Oil Spill, Land, Look-alike	96.05	51.60	55.60	52.55	91.81	69.52
Ship, Land, Oil Spill, Look-alike	96.05	51.60	55.60	52.55	91.81	69.52
Land, Ship, Oil Spill, Look-alike	96.05	51.60	55.60	52.55	91.81	69.52

training times. However, training YOLOv8 and the Adapter is more efficient than training the entire SAM. A lighter SAM version, like EfficientSAM [15], could improve SAM-OIL's training and inference speed.

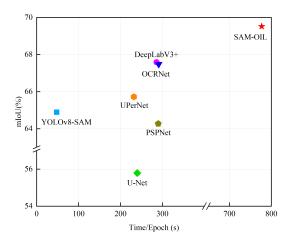


Fig. 2. The epoch time for each model.

## C. Ablation Study

1) Detector accuracy: To assess the impact of detector performance on the accuracy of SAM-OIL, ground truth

bounding boxes (gt\_box) are input into SAM and Adapted SAM respectively. In Table II, YOLOv8-SAM (gt\_box) surpasses YOLOv8-SAM by 12.2%, and SAM-OIL (gt\_box) surpasses SAM-OIL by 10.83%. It demonstrates that the detector significantly impacts the accuracy within the SAM-OIL framework.

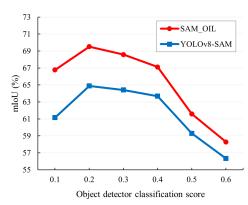


Fig. 3. The influence of object detector classification scores on the accuracy.

2) Detector classification scores: Besides bounding boxes and their corresponding categories, the detector also outputs classification scores, which can help determine if boxes should be ignored. Fig. 3 shows that different score thresholds impact accuracy; with a 0.2 score, YOLO-SAM and SAM-OIL reach

peak accuracy. Increasing the score significantly lowers accuracy as higher scores may filter out correct detections.

- 3) OMF: Table III illustrates that YOLOv8-SAM+OMF outperforms YOLOv8-SAM by 0.93%, and SAM-OIL exceeds YOLOv8-SAM+Adapter by 1.55%. Notably, OMF operates without training parameters. YOLOv8-SAM+OMF reaches 65.82% mIoU and needs just box-level annotated samples for training, proving the potential of SAM-based composite methods in reducing the workload of sample annotation.
- 4) Adapter: In Table III, YOLOv8-SAM+Adapter surpasses YOLOv8-SAM by 3.08%, and SAM-OIL surpasses YOLOv8-SAM+OMF by 3.7%. Furthermore, comparing YOLOv8-SAM in Table I with Random in Table IV shows that the Adapter module increases IoU by 6.5% for "oil spill" and 6.72% for "look-alike". This emphasizes the role of Adapter in enhancing SAM's ability to segment objects with ambiguous boundaries.
- 5) Order of OMF: SAM outputs masks in four categories: ship, land, oil spill, and look-alike, with 24 possible permutations plus a random order, totaling 25. Table IV shows accuracy for random, worst, and best orders, emphasizing a 3.24% accuracy gap between the best and worst. This suggests order significantly impacts accuracy, advising higher priority for "ship" and "land" due to their effect on classification accuracy and size in images, and lower for "look-alike" due to its proximity to other categories and lower accuracy, to improve overall performance.

## D. Qualitative Analysis

Fig. 4 shows that YOLOv8-SAM and SAM-OIL perform well in detecting elongated and small dark areas with simple surroundings, as seen in column 1. The integration of OMF and Adapter helps SAM-OIL segment objects with fuzzy boundaries more accurately, as shown in columns 2 and 3. However, small-sized look-alikes, particularly in the top left of column 3 and the bottom left of column 4, are sometimes misclassified as oil spills by the SAM-OIL, highlighting the challenges of using SAR imagery alone for oil spill detection.

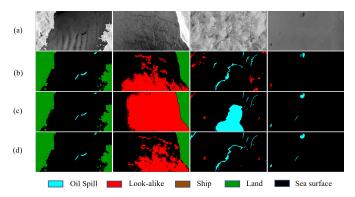


Fig. 4. Qualitative examples of YOLOv8-SAM and SAM-OIL. (a) SAR Images, (b) Ground Truth, (c) YOLOv8-SAM, (d) SAM-OIL.

## IV. CONCLUSION

We proposed a composite framework, SAM-OIL, which integrates YOLOv8 with the SAM, incorporating OMF and

Adapter modules. To the best of our knowledge, this is the first time the SAM has been applied in oil spill detection. Compared with existing methods, SAM-OIL achieves the highest mIoU on the M4D dataset with 69.52%, demonstrating the significant potential of the SAM-based composite framework in oil spill detection. Furthermore, the SAM-OIL version without the adapter module, YOLOv8-SAM+OMF, achieved 65.82% mIoU, indicating the framework's ability on reducing the dependence on finely annotated samples.

Qualitative analysis shows that small-sized look-alikes are often confused with oil spills, suggesting the limitations in relying solely on SAR satellite imagery for detection and the need of ancillary data (e.g., oil platform locations, ship traffic) [3]. Within the SAM-OIL framework, the object detector significantly contributes to the accuracy improvement. Hence, prioritizing object detectors could further improve detection performance in the future.

#### REFERENCES

- [1] C. Brekke and A. H. Solberg, "Oil spill detection by satellite remote sensing," *Remote Sens. Environ.*, vol. 95, no. 1, pp. 1–13, 2005.
- [2] R. Al-Ruzouq, M. B. A. Gibril, A. Shanableh, A. Kais, O. Hamed, S. Al-Mansoori, and M. A. Khalil, "Sensors, features, and machine learning for oil spill detection and monitoring: A review," *Remote Sens.*, vol. 12, no. 20, p. 3338, 2020.
- [3] W. Alpers, B. Holt, and K. Zeng, "Oil spill detection by imaging radars: Challenges and pitfalls," *Remote Sens. Environ.*, vol. 201, pp. 133–147, 2017.
- [4] Y. Dong, Y. Liu, C. Hu, I. R. MacDonald, and Y. Lu, "Chronic oiling in global oceans," *Science*, vol. 376, no. 6599, pp. 1300–1304, 2022.
- [5] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and I. Kompatsiaris, "Oil spill identification from satellite images using deep neural networks," *Remote Sens.*, vol. 11, no. 15, p. 1762, 2019.
- [6] C. Brekke and A. H. Solberg, "Classifiers and confidence estimation for oil spill detection in envisat asar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 65–69, 2008.
- [7] Q. Zhu, Y. Zhang, Z. Li, X. Yan, Q. Guan, Y. Zhong, L. Zhang, and D. Li, "Oil spill contextual and boundary-supervised detection network based on marine sar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, 2021.
- [8] R. Hasimoto-Beltran, M. Canul-Ku, G. M. D. Méndez, F. J. Ocampo-Torres, and B. Esquivel-Trava, "Ocean oil spill detection from sar images based on multi-channel deep learning semantic segmentation," *Mar. Pollut. Bull.*, vol. 188, p. 114651, 2023.
- [9] X. Ma, J. Xu, P. Wu, and P. Kong, "Oil spill detection based on deep convolutional neural networks using polarimetric scattering information from sentinel-1 sar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021.
- [10] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, October 2023, pp. 4015–4026.
- [11] K. Chen, C. Liu, H. Chen, H. Zhang, W. Li, Z. Zou, and Z. Shi, "Rsprompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–17, 2024.
- [12] M. Contributors, "MMYOLO: OpenMMLab YOLO series toolbox and benchmark," https://github.com/open-mmlab/mmyolo, 2022.
- [13] L. Ke, M. Ye, M. Danelljan, Y. Liu, Y.-W. Tai, C.-K. Tang, and F. Yu, "Segment anything in high quality," in *Proc. Adv. Neural Inf. Process.* Syst., vol. 36, 2023.
- [14] M. Contributors, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark," https://github.com/open-mmlab/ mmsegmentation, 2020.
- [15] Y. Xiong, B. Varadarajan, L. Wu, X. Xiang, F. Xiao, C. Zhu, X. Dai, D. Wang, F. Sun, F. Iandola *et al.*, "Efficientsam: Leveraged masked image pretraining for efficient segment anything," *arXiv* preprint arXiv:2312.00863, 2023.