

SAR Semantic Segmentation

Batyr Bodaubay, Bekzat Ongdassynov, Diana Ulaskhanova, Daria Gole, Temirlan Zhexembeyev

I. INTRODUCTION

Marine oil pollution has consistently posed a significant threat to the marine environment. Oil released into the ocean spreads with the wind and currents, continuing to move across the ocean surface and leading to widespread pollution. Hence, rapid detection of marine oil spills is crucial for prompt response and reducing environmental harm. Synthetic Aperture Radar (SAR) has been widely used for oil spill detection in recent years. SAR is an active sensor which transmits electromagnetic waves to the Earth's surface and receives reflected signals to generate images by synthesizing the multiple aperture echoes [1]. SAR-based oil spill detection is often hindered by the presence of look-alikes. These substances can be produced by wind, rain, ocean upwelling, and biological processes and have backscattering coefficients and morphological characteristics that are similar to those of actual oil slicks. As a result, SAR-based image segmentation can often lead to false detections, posing a significant challenge in accurately identifying oil spills. To differentiate oil slicks from look-alikes in SAR images, several strategies have been developed, including multi-class segmentation, modifications to CNN architectures, and analyses of image textures [1].

II. RELATED WORK

In the realm of detecting and addressing oil spills through image segmentation, recent studies have showcased various approaches and models to tackle this critical environmental challenge. Ahmed, S. et al. describe a custom data generator that utilizes a modified Segmentation Network (Seg-Net) within a Conditional Generative Adversarial Network (CGAN) framework. The proposed model shows an average accuracy of 99.04, an Intersection over Union (IoU) index of 96.59 percent, and a precision of 85.24 percent. [2]. Yang, Y. J. et al. proposed YOLOv4 objection detection model for oil spills in the Eastern Mediterranean Sea. The trained object detector has an average precision of 69.10 percent and 68.69 percent on the validation and test sets, respectively [3]. Krestenitis et al. propose deep convolutional neural networks (DCNNs) as an effective method. The dataset provided is used to evaluate the performance of established DCNN segmentation models for this specific task. The model accomplished 65.04 percent accuracy on the test set [4].

III. METHODOLOGY

General idea of our work is to use instance segmentation capabilities of YOLOv8 model to perform semantic segmentation on the oil spill dataset.

A. Dataset

Oil Spill Detection Dataset contains jpg images extracted from satellite Synthetic Aperture Radar (SAR) data depicting oil spills and other relevant instances, as well as their corresponding ground truth masks. The developed dataset (400MB) contains around 1000 images for training and 110 images for testing, depicting instances of 5 classes, namely oil spill, look-alike, land, ship and sea areas.

B. Data Preprocessing

Apart from the preprocessing in the original dataset we had perform data transformation to the suitable format so that YOLO models could train.

C. Selecting Models

1) *U-Net*: In our segmentation work, we employed a U-Net model and focused on binary classification, specifically identifying and segmenting areas corresponding to oil spills within the images [5]. This simplifies the task to delineating between 'oil spill' and 'non-oil spill' regions.

The encoder component is responsible for feature extraction and downsampling the input image. In our implementation, we utilized a ResNet18 backbone as the encoder. In the decoder section, the extracted features from the encoder are gradually upsampled to reconstruct the segmentation mask. This process involves a series of deconvolutional layers to increase the spatial resolution of the feature maps. Skip connections concatenate feature maps from the encoder with corresponding feature maps in the decoder. This allows the model to retain fine-grained details from the early stages of the encoding process, aiding in precise localization during segmentation.

The total number of parameters in our U-Net model amounts to 24,456,154. Among these, 24,438,804 are trainable parameters, while 17,350 are non-trainable parameters.

2) *YOLOv8*: Our multi-class detection task leverages the YOLOv8 model, a cutting-edge, single-shot detector renowned for its speed and accuracy. YOLOv8 operates on a grid-based prediction principle, dividing the input image into a grid system and performing object detection within each grid cell. This approach, coupled with anchor boxes, allows the model to efficiently predict

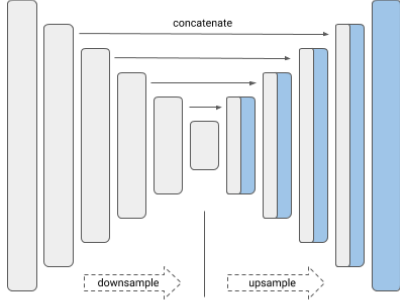


Figure 1: U-Net model simple architecture

bounding boxes and classify objects simultaneously. We chose YOLOv8 due to its inherent advantages over traditional two-stage detectors. While two-stage models involve separate region proposal and classification steps, YOLOv8 streamlines the process with a single convolutional neural network. This unified architecture leads to faster training times. Specifically, our implementation utilizes the YOLOv8n-seg model variant, a robust architecture comprising 261 layers and 3,409,968 parameters. To prepare our dataset, we converted the training masks into YOLO's text file format, ensuring compatibility and efficient data loading during the training process. This conversion process involved encoding the bounding box coordinates and class labels for each object present in the training images.

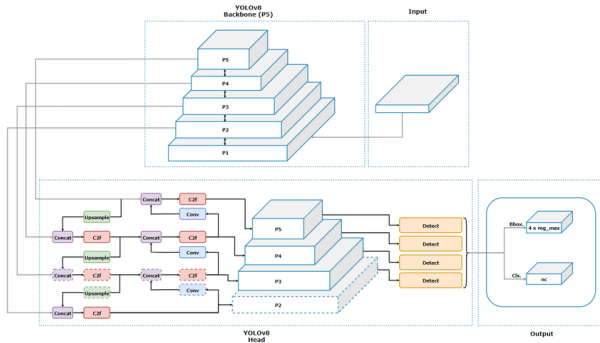


Figure 2: YOLOv8 simple architecture

D. Evaluation methods

To assess the performance of the trained model, we employed the Intersection over Union (IoU) metric, a common evaluation metric used in object detection and segmentation tasks. IoU quantitatively evaluates how close the predicted segmentation masks are to the ground truth masks. The formula for IoU is given as follows:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

This can be further expressed in mathematical notation:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

Where:

- A represents the set of pixels belonging to the predicted mask.
- B represents the set of pixels belonging to the ground truth mask.

IoU Calculation Procedure:

To compute the IoU, the following steps were followed for each test image:

- 1) Both the predicted mask and the ground truth mask were converted into binary format where pixels of interest are marked as 1 and all others as 0.
- 2) The intersection ($A \cap B$) was determined by taking the pixel-wise logical AND between the predicted and ground truth masks.
- 3) The union ($A \cup B$) was determined by taking the pixel-wise logical OR between the predicted and ground truth masks.
- 4) The IoU was then calculated using the formula provided above.

Per-Class IoU:

Additionally, we computed the IoU for each class separately to analyze the model's performance across different types of objects. This involved isolating the mask for each class before applying the IoU calculation formula.

This method allowed us to quantitatively evaluate the segmentation accuracy of our model, providing a clear measure of how effectively the model can distinguish between different classes and how closely it matches the actual object boundaries.

IV. RESULTS

The table I provides a comparative analysis of the two models, YOLOv8 outperforms U-Net, however low U-Net score relative to literature available suggests that the setup is subject to further analysis. It is worth noting that the backbone for the model, loss functions, and other important points of model specification are very similar across our research and the PAPER, thus such low scores are probably attributed to some other reason, no further adequate comparison can be made due to closed source nature of the literature.

Model	IoU score
U-Net	0.12
YOLOv8	0.232

Table I: IoU scores for oil spill class

The dataset for YOLOv8 underwent no additional preprocessing other than changes applied on images to be

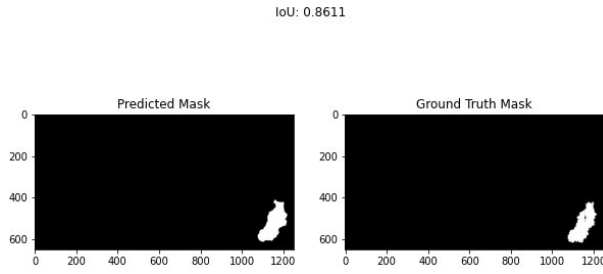


Figure 3: YOLOv8 prediction with a high score

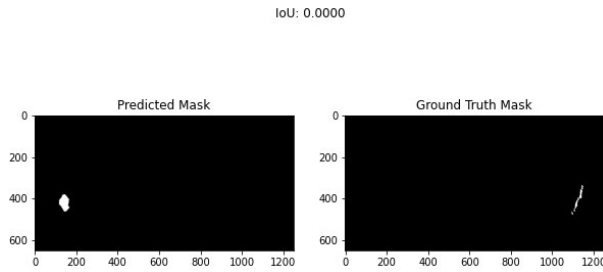


Figure 4: YOLOv8 prediction with a low score

properly fed to the model. It suggests that the difference in IoU score is mostly contributed to how models then classified oil spills.

Some observations:

Despite the low overall IoU scores, we can see in Figure 3 that some predictions were fairly accurate, though we also see predictions like in Figure 4 that not only fail to recognize oil spill, but also misclassify other objects as our target class.

Figures 5 and 6 show how U-Net creates predictions. It can be seen that on top having the same issue as YOLOv8 with lookalikes, U-Net also misses the general shape of the object.

V. CONCLUSION

With references to the study from 'Remote Sensing' [4], we advise readers to visit and see tables with IoU results for better context.

As noted in results, scores were lower than expected which could be attributed to several reasons:

- Closed source nature of literature at hand. It could be the case that some stages of classification benefited from decisions that are not reflected in papers.
- Limited training times. Our models could still show better results given better parameter choice and more epochs in training. However, as noted before, the choices were very similar to those made by other authors. During training, we observed how models converged, thus, have reached reasonable readings.
- Limitations of the dataset. SAR images were of basic RGB channels and have been preprocessed to greyscale, which is a potential loss in information essential for models.

Upon observation of label and prediction masks next to corresponding preprocessed images, it became clear that the problem lies mostly in dataset not being complex enough for models and having rather simple features. Usually SAR images for such tasks include other channels that give deeper insight and our dataset provides very little for distinguishing between actual oil spills and other objects. The shapes may be fairly similar, and colors intersect. The problem is also apparent when we compare studies that use different datasets with more to offer for model learning - they surpass the upper limit of around 0.7 IoU. Although cannot be measured and accurately pinpointed, the difference seems to closely relate to dataset limitations.

REFERENCES

- [1] C. Li, D. jin Kim, S. Park, J. Kim, and J. Song, "A self-evolving deep learning algorithm for automatic oil spill detection in sentinel-1 sar images," *Remote Sensing of Environment*, vol. 299, p. 113872, 2023.
- [2] S. Ahmed, T. ElGharbawi, M. Salah, and M. El-Mewafi, "Deep neural network for oil spill detection using sentinel-1 data: application to egyptian coastal regions," *Geomatics, Natural Hazards and Risk*, vol. 14, no. 1, pp. 76–94, 2023.
- [3] Y. J. Yang, S. Singha, and R. Mayerle, "A deep learning based oil spill detector using sentinel-1 sar imagery," *International Journal of Remote Sensing*, vol. 43, no. 11, pp. 4287–4314, 2022.
- [4] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Aygerinakis, S. Vrochidis, and I. Kompatsiaris, "Oil spill identification from satellite images using deep neural networks," *Remote Sensing*, vol. 11, no. 15, p. 1762, 2019.
- [5] P. Iakubovskii, "Segmentation models." https://github.com/qubvel/segmentation_models, 2019.



Figure 5: U-Net prediction



Figure 6: U-Net prediction