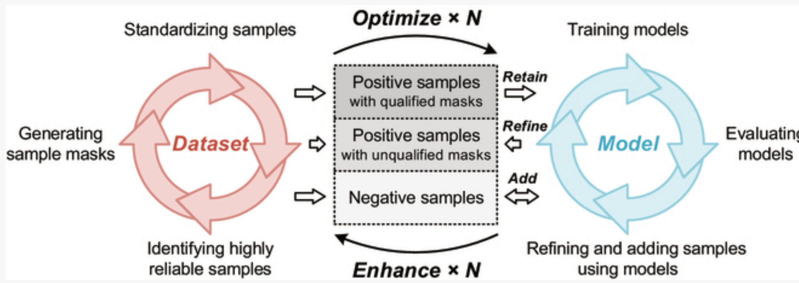
New dataset and AI model transform ocean pollution tracking

GA, UNITED STATES, June 30, 2025 /EINPresswire.com/ -- Researchers have developed the largest and most comprehensive global dataset for oil slick imagery and an advanced deep learning model to detect oil spills from Sentinel-1 satellite data automatically. Their method significantly improves segmentation accuracy under real-world conditions, addressing key challenges in marine pollution monitoring. The approach supports faster response and regulation, heralding a new era for satellite-based environmental protection.



Snowballing framework between the oil slick dataset and the deep learning model.

Oil slicks from human activities and natural seeps pose severe threats to marine ecosystems, public health, and coastal economies. Satellite-based synthetic aperture radar (SAR) systems offer all-weather, global-scale monitoring; however, detecting oil slicks remains challenging due to their diverse appearances and confusion with similar features, such as biogenic slicks and low-wind areas. Manual interpretation is time-consuming and prone to errors, while existing automatic methods lack robustness, adaptability, and generalizability. The absence of large, diverse training datasets further limits the potential of deep learning (DL) models. Due to these challenges, there is an urgent need to develop a scalable, high-precision approach for segmenting global oil slicks.

A research team from Nanjing University has introduced a deep learning framework to enhance global oil slick detection using Sentinel-1 SAR imagery. The study published in May 2025 in Journal of Remote Sensing, tackles longstanding challenges in oil spill segmentation by combining a massive annotated dataset with a novel model training strategy. By addressing data scarcity and model generalization issues, the work offers a powerful, operational-grade tool for ocean monitoring and environmental response.

The team created the GlobalOSD-SAR dataset—containing over 100,000 annotated oil slick and lookalike images—spanning global offshore regions and varying environmental conditions. Leveraging this dataset, they trained a deep learning model using the UNet++ architecture with a unique co-training and "snowballing" framework that progressively refines data quality. The optimized model achieved outstanding performance in real-world conditions, reaching Intersection over Union (IoU) scores of 82.63–95.99%. These results mark a substantial leap in segmentation accuracy compared to earlier methods and demonstrate that enriching dataset diversity can significantly boost model reliability and generalization.

To construct GlobalOSD-SAR, researchers analyzed over half a million Sentinel-1 SAR images, selecting high-confidence oil slicks based on recurrence, ancillary evidence (e.g., ship routes, optical imagery), and visual verification. Negative samples representing lookalike phenomena were similarly identified and enriched using model-generated false positives. An automatic threshold-based algorithm and manual inspection were employed to ensure the creation of high-quality pixel-level masks.

For model training, the team employed UNet++ with pre-trained ImageNet weights. They introduced a co-training strategy that merged oil slick and lookalike samples to improve generalization and iteratively refined training data through a "snowballing" framework that enhanced both positive and negative sample quality.

Experiments revealed that increasing the diversity of training samples—across regions, categories, wind conditions, and incidence angles—substantially improved model performance. Compared to traditional training methods, the model using co-training and the snowballing framework consistently achieved lower dice loss and higher IoU scores, even under challenging test scenarios.

"Our model's real-world performance proves that a comprehensive dataset and iterative learning can overcome the limitations of earlier approaches," said Prof. Yongxue Liu, corresponding author. "This work not only enhances oil spill monitoring but also sets a foundation for broader, AI-driven marine pollution surveillance using satellite imagery."

The study utilized Sentinel-1 GRD SAR imagery from 2014 to 2021, focusing on the VV polarization band, which is optimal for oil slick detection. Positive samples were verified using ancillary data, while negative samples were expanded through iterative correction of false positives. Image masks were generated using a dual-thresholding and spatial filtering algorithm. Standardized 1,024 × 1,024-pixel image patches were used for model training. The model architecture, UNet++, was fine-tuned with learning rate control, grayscale normalization, and co-training strategies to maximize performance across diverse conditions.

The researchers plan to scale up their model by incorporating high-resolution texture metrics and Transformer-based architectures to improve accuracy in complex environments. They also aim to expand coverage in underrepresented regions, such as high-latitude seas, and integrate real-time monitoring capabilities. The GlobalOSD-SAR dataset is now available for public use, enabling the broader scientific community to develop, test, and benchmark advanced algorithms for marine oil pollution detection, thereby supporting faster, data-informed, and more resilient environmental governance.

References
DOI
10.34133/remotesensing.0613

Original Source URL
https://doi.org/10.34133/remotesensing.0613

Funding information
This research was sponsored by the Key Research and Development Program of China (Grant no. 2019YFA0606601) and the Fundamental Research Funds for the Central Universities (0904-14380035).

Lucy Wang
BioDesign Research
email us here
