Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

# Enhancing Offshore Infrastructure Monitoring: Synthetic Data Generation for Deep Learning-Based Object Detection on Sentinel-1 Radar Imagery

Robin Spanier[1], Thorsten Höser[1], Marco Ottinger[1] and Claudia Künzer[1,2]

[1] German Aerospace Center (DLR), Earth Observation Center (EOC), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, 82234 Wessling, Germany
[2] University of Wuerzburg, Department of Remote Sensing, Wuerzburg, Germany

## Monitoring offshore infrastructure expansion

The recent and ongoing expansion of marine infrastructure, including offshore wind farms, oil and gas platforms, artificial islands, and aquaculture facilities, highlights the need for effective monitoring systems. Precise quantification in space and time is crucial to planning the future expansion, usage, management, and impact of marine offshore infrastructure [1].

We applied the object detector YOLOv10 to efficiently detect and classify offshore infrastructure objects (specifically offshore oil and gas platforms) on Sentinel-1 radar imagery in three diverse test regions: the Gulf of Mexico, the North Sea, and the Persian Gulf.

Fig. 1: Complexity of platform occurances. Google Earth, Sentinel-1 Radar Image and aerial photography of a single platform, Mittelplate in Germany and two different platform clusters, Zakum West Supercomplex and Upper Zakum production facility in Abu Dhabi..

## Importance of synthetic training data in the remote sensing context

The development of a robust and reliable object detection model depends on the availability of comprehensive, balanced training datasets. For our training dataset we hand labeled the South China Sea, Caspian Sea, Gulf of Guinea and Coast of Brazil (table 1). Manual annotation of existing objects is the standard method for dataset creation, but it falls short specifically in remote sensing when samples in the real world are scarce, particularly for underrepresented object classes, shapes, and sizes like the platform clusters in our dataset (fig.1).

To address this limitation, we generated our own synthetic training images in an object-oriented approach, assembling individual object instances into coherent and meaningful remote sensing scenes (fig 2).

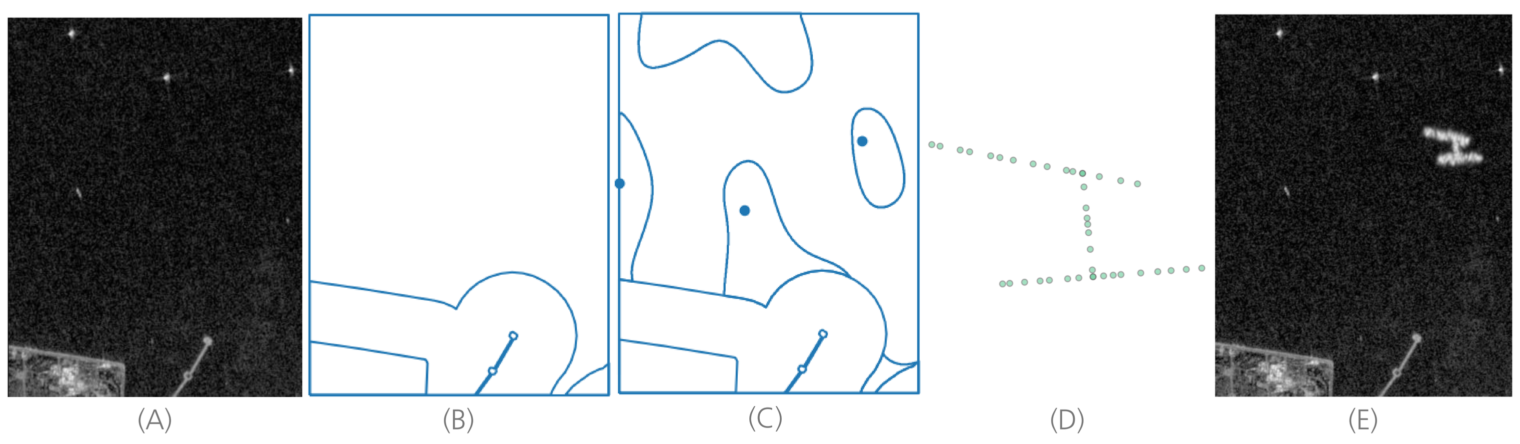


Fig. 2: Simplyfied workflow of generating synthetic training samples with (A) the input scene, (B) generation of a mask with a 1000m buffer from the coast, (C) randomly placed potential point locations inside Perlin Noise polygons, (D) generation of complex and randomised geometries imitating platform cluster structures, (E) insertion of the geometry hull into one of the potential locations into the scene with a KDE2D kernel to produce texture.

## Offshore platform detection with YOLOv10 and synthetic training data

We prepared two different training datasets. For our base dataset we hand labeled the South China Sea, Caspian Sea, Gulf of Guinea and Coast of Brazil. For the second training dataset we added synthetic training samples of platform clusters (fig. 3).

We compiled a validation ground truth dataset for three different and diverse test sites: the Gulf of Mexico, the North Sea, and the Persian Gulf to test our two models.

Table 1: Ground Truth datasets.

| dataset | purpose | labels | | | |
|---|---|---|---|---|---|
| | | platform | platform cluster | windturbine | ∑ |
| North Sea | real world data hand labeled | 425 | 80 | 4950 | 5455 |
| Persian Gulf | | 1538 | 162 | | 1700 |
| Gulf of Mexico | | 1554 | 168 | | 1722 |
| South China Sea | | 1054 | 98 | 2920 | 4072 |
| Caspian Sea | | 481 | 133 | | 614 |
| Gulf of Guinea | | 932 | 89 | | 1021 |
| Coast of Brazil | | 33 | 3 | | 36 |
| Synthetic | for class balance | | 248 | | |
| NS-PG-GoM | validation | 3517 | 410 | 4950 | 8877 |
| SCS-CS-GoG-CoB | base training dataset | 2500 | 323 | 2920 | 5743 |
| SCS-CS-GoG-CoB + synthetic | base training dataset + synthetic | 2500 | 571 | 2920 | 5991 |

Table 2: Comparison of detection results of the two different trainings.

| | overall acc. | avg pr | avg rc | avg F1 | platform cluster | | |
|---|---|---|---|---|---|---|---|
| | | | | | pr | rc | F1 |
| Base dataset | 0.85 | 0.91 | 0.91 | 0.91 | 0.78 | 0.77 | 0.78 |
| Base dataset + synthetic | 0.87 | 0.92 | 0.90 | 0.91 | 0.82 | 0.74 | 0.78 |

## Inference results

The inference results (table 2) show the impact of our synthetic data generation approach on the training and how unbalanced classes can be better represented and model performance improved. This underscores the critical importance of balanced datasets and highlights synthetic data generation as an effective strategy to address common challenges in remote sensing.

Figure 3 shows all platforms (purple) and platform clusters (green) that were observed in the North Sea (A), Persian Gulf (B) and Gulf of Mexico (C).

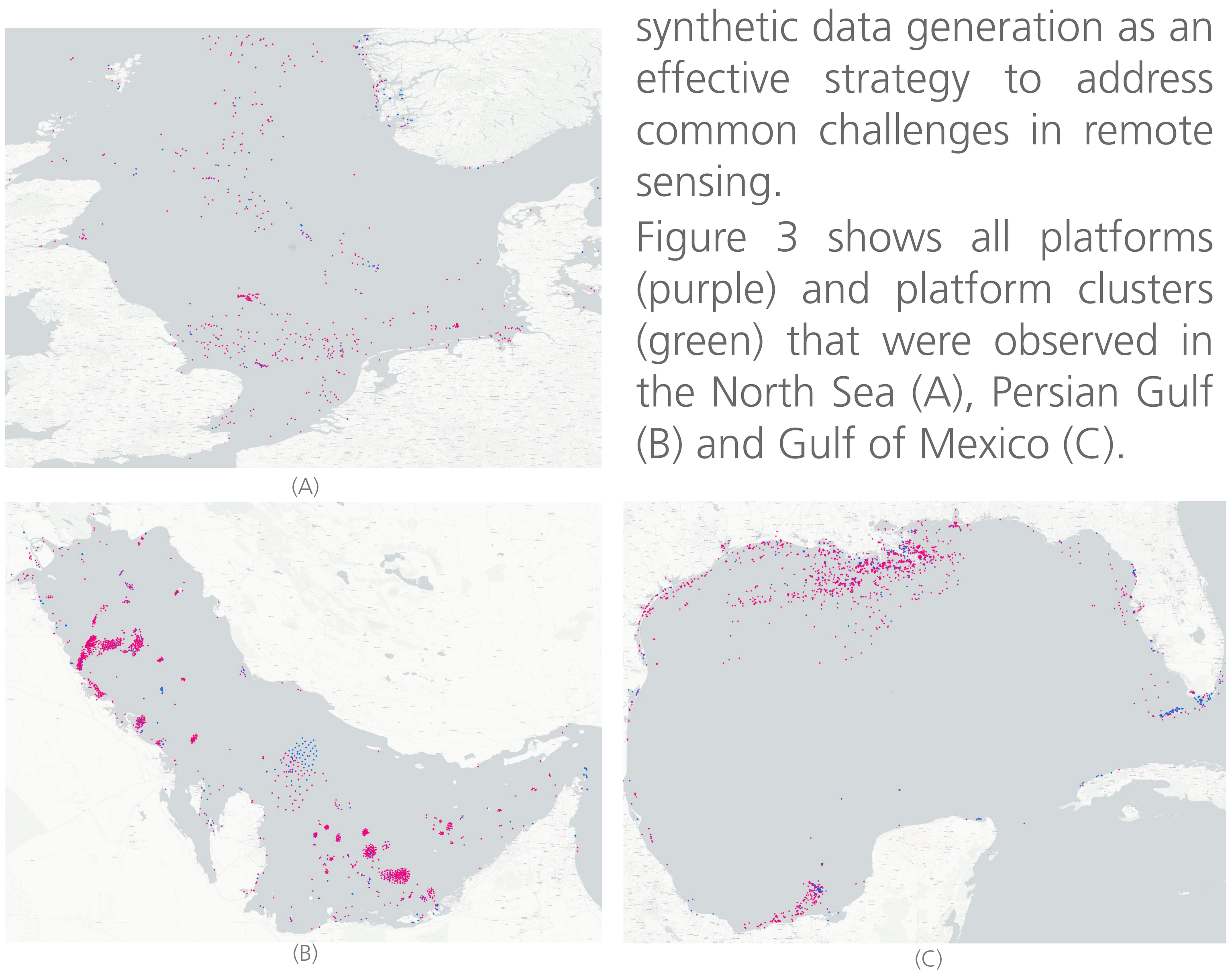


Fig. 3: Detection results of the base training dataset enhanced by synthetic training data for the three regions North Sea, Gulf of Mexico and Persian Gulf (platforms: pink, clusters: blue).

[1] Spanier, R., Kuenzer, C., 2024:Marine Infrastructure Detection with Satellite Data—A Review. *Remote Sens.*, 16, 1675. https://doi.org/10.3390/rs16101675

DLR
powered by terrabyte

Robin Spanier
German Aerospace Center (DLR)
Earth Observation Center (EOC)
German Remote Sensing Data Center (DFD)

Oberpfaffenhofen
82234 Weßling
Tel.: +49 8153 28-1006
robin.spanier@dlr.de