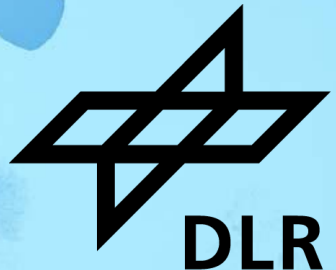


FUSION OF OIL SPILL DETECTION RESULTS FROM THRESHOLDING AND DEEP LEARNING USING LANDSAT DATA OVER THE NORTH SEA

Olga Schmidt, Carolin Wloczyk, Egbert Schwarz, Detmar Krause

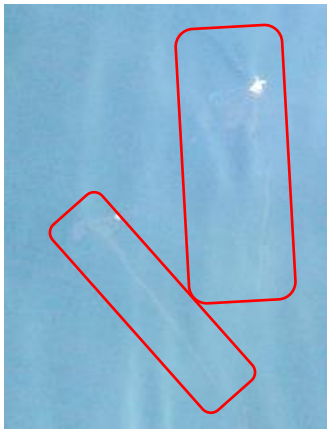
**Maritime Safety and Security Lab Neustrelitz, German Aerospace Center (DLR),
Germany**



Introduction

- Risks of oil spills: extraction, transport and use of oil
→ Danger to human health & major impact on the marine and coastal environment
- Monitoring of large areas and timely and accurate oil spill detection using **remote sensing**
- Highly variable appearance of oil spills on **optical satellite images**
- Darker or brighter than surrounding water
- Contrast is influenced by oil properties (thickness, type), surface conditions, clouds, haze, fog, sun glint

Oil spills on optical images (Landsat):



Thresholding Method

Specifically developed for oil spill detection in the North Sea

Input data

- Landsat 8/9 OLI Level-1 data

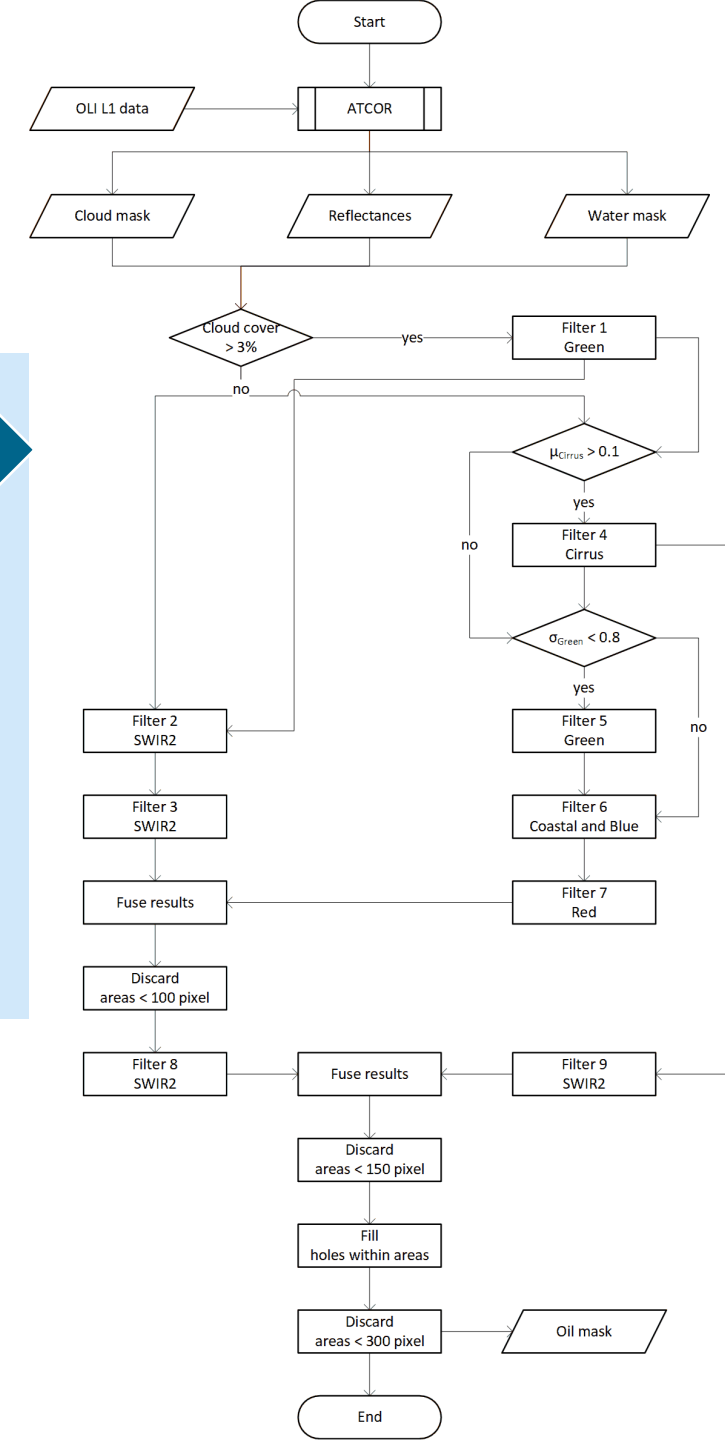
Pre-processing

- Atmospheric correction using ATCOR model¹
- Extraction of cloud and water classes from ATCOR output
- Water masking: Oil is included in water class → water mask defines the search domain

Thresholding process

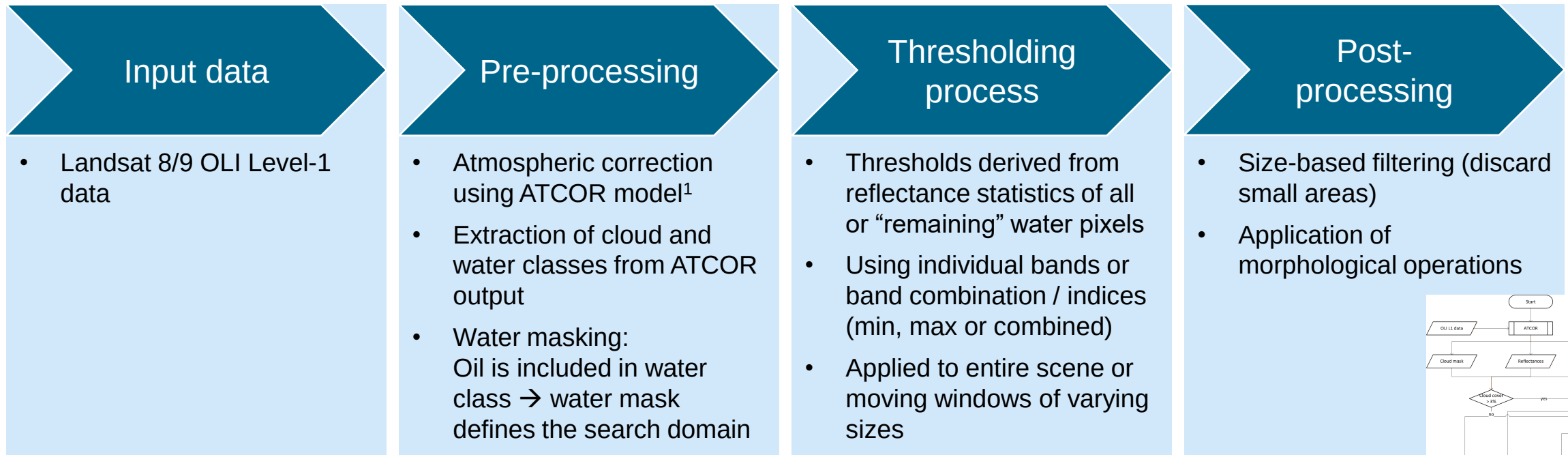
- Thresholds derived from reflectance statistics of all or “remaining” water pixels
- Using individual bands or band combination / indices (min, max or combined)
- Applied to entire scene or moving windows of varying sizes

¹ R. Richter and D. Schlöpfer, “Atmospheric / Topographic Correction for Satellite Imagery,” ATCOR-2/3 User Guide, Version 9.5, 2023.

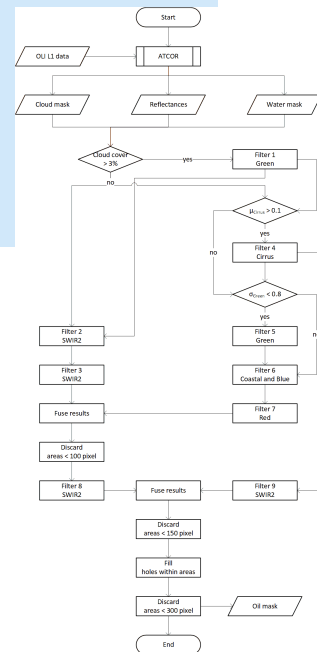


Thresholding Method

Specifically developed for oil spill detection in the North Sea



¹ R. Richter and D. Schlöpfer, “Atmospheric / Topographic Correction for Satellite Imagery,” ATCOR-2/3 User Guide, Version 9.5, 2023.



Deep Learning Approach



Input data

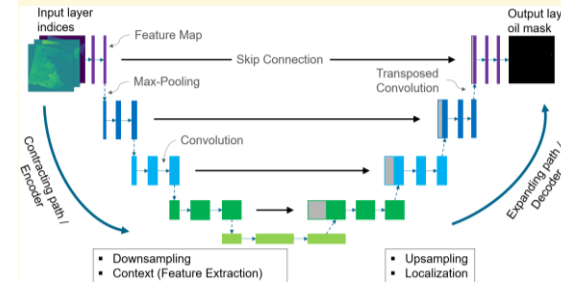
- Landsat 8/9 OLI Level-1 data

Pre-processing

- Atmospheric correction using ATCOR model¹
- Index calculation
- Splitting into patches (128x128 px)

Inference / prediction

- U-Net model with ResNet-50 backbone (92 training patches)
- Model inference with probability > 0.7



	Precision	Recall	F1-Score
0	0.95	1.00	0.97
1	0.94	0.87	0.90

Post-processing

- Merging patches
- Cloud / cloud shadow masking^{2,3}
- Size-based filtering
- Morphological operations

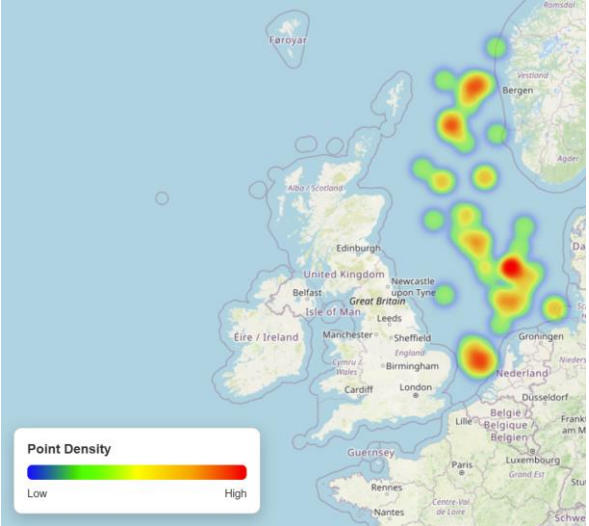
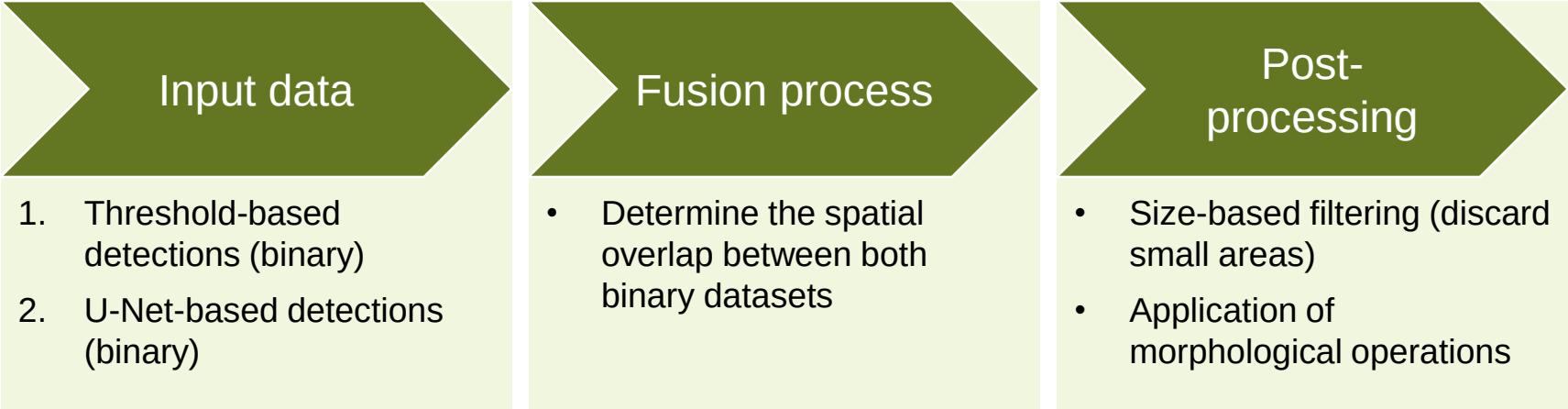
¹ R. Richter and D. Schl pfer, "Atmospheric / Topographic Correction for Satellite Imagery," ATCOR-2/3 User Guide, Version 9.5, 2023.

² UKIS Cloud Shadow MASK (ukis-csmask), <https://github.com/dlr-eoc/ukis-csmask?tab=readme-ov-file>

³ Learning Attention Network Algorithm (LANA), [GitHub - hankui/Cloud-mask-for-Landsat-8-9](https://github.com/hankui/Cloud-mask-for-Landsat-8-9): LANA cloud mask codes for Landsat 8/9 · GitHub



Fusion - Methodology

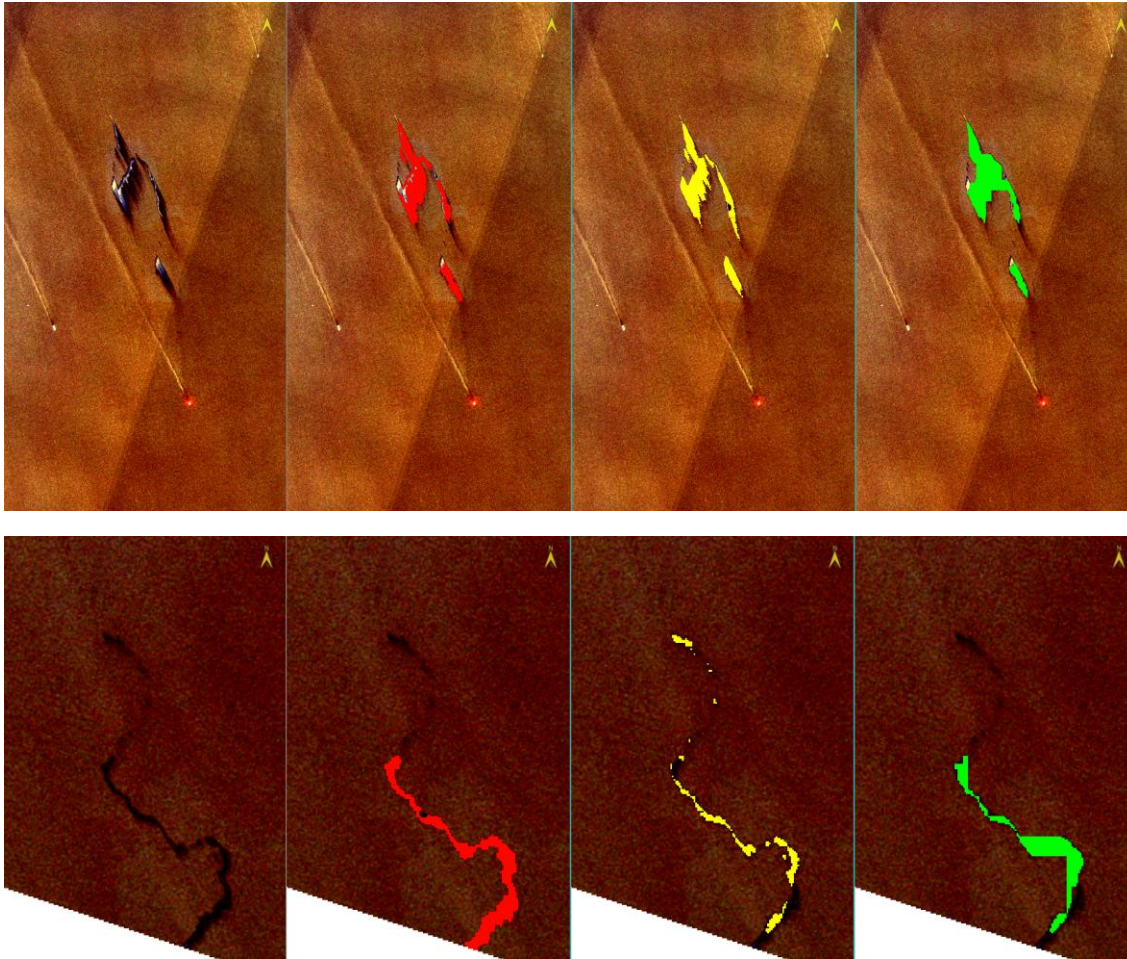


➔ **Hypothesis:** U-Net and thresholding produce different false positive detections.

Expectation: Fusion of both methods leads to a reduction in false positive detections and improves detection accuracy.

Data basis for the fusion process	
Area	North Sea
Sensor	Landsat-8
Time period	2013 – 2025
Number of scenes processed	48
Number of processed scenes with an available binary oil mask	15

Fusion - Results



-  Thresholding based detections
-  U-Net detections
-  Fusion output (after post-processing)

False positive reduction

Thresholding: 86.9 %
U-Net: - 13.7 %

Evaluation metrics (fusion)

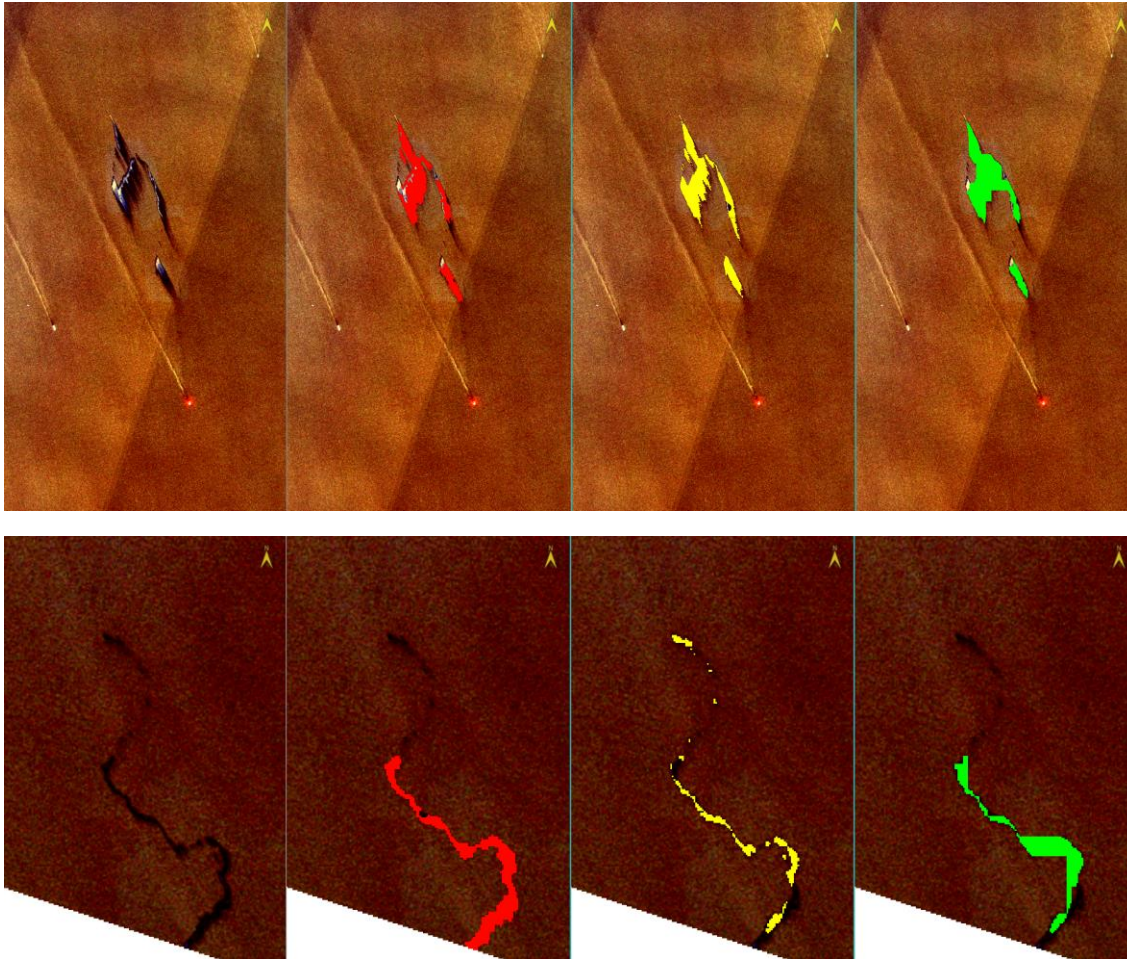
Recall: 0.7226
Precision: 0.9856
F1-Score: 0.8339

- 2019-08-26
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

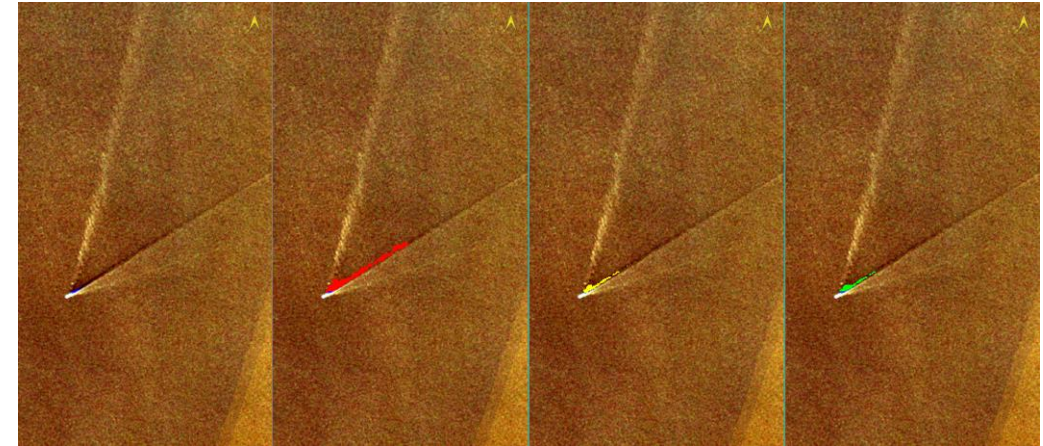


Good performance while preserving most oil pixels.

Fusion - Results



Remaining false positive detections after fusion:



-  Thresholding based detections
-  U-Net detections
-  Fusion output (after post-processing)

False positive reduction

Thresholding: 86.9 %
U-Net: - 13.7 %

Evaluation metrics (fusion)

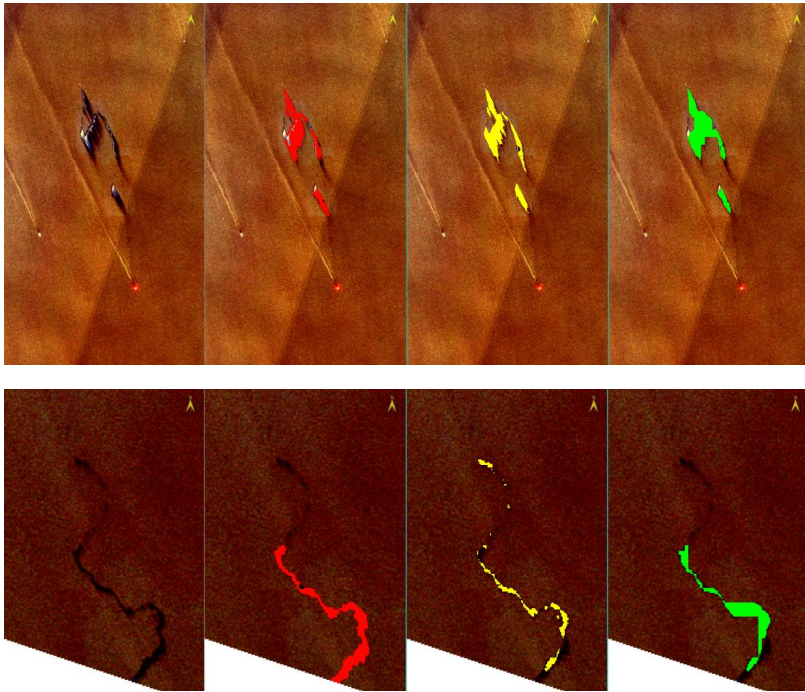
Recall: 0.7226
Precision: 0.9856
F1-Score: 0.8339

- 2019-08-26
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

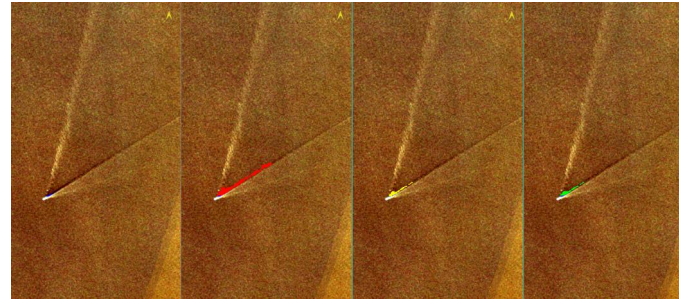


Good performance while preserving most oil pixels.

Fusion - Results



Remaining false positive detections after fusion:



- Thresholding based detections
- U-Net detections
- Fusion output (after post-processing)

False positive reduction

Thresholding: 86.9 %
U-Net: - 13.7 %

Evaluation metrics (fusion)

Recall: 0.7226
Precision: 0.9856
F1-Score: 0.8339

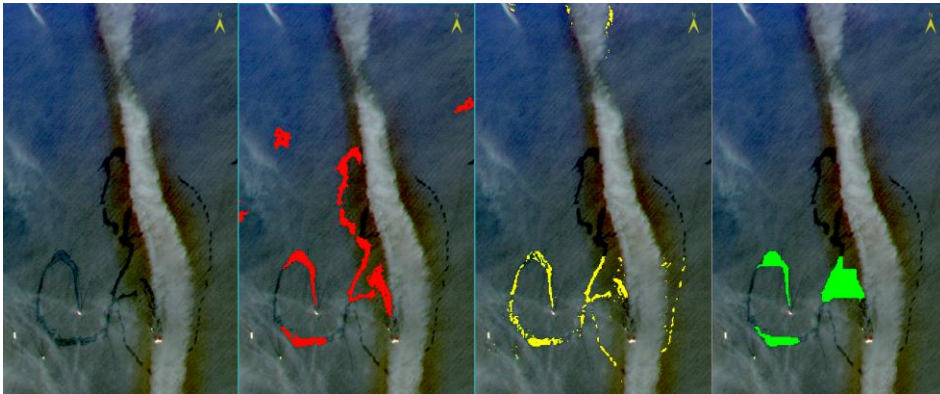
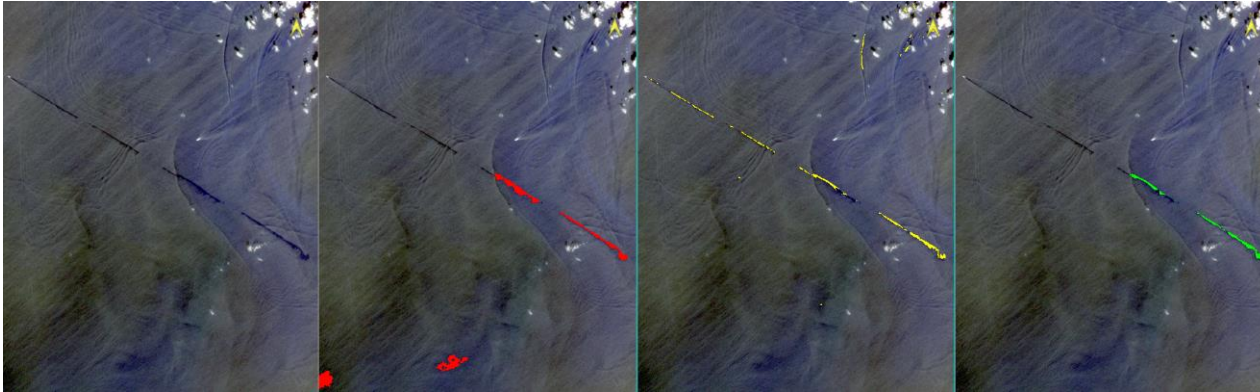
Evaluation based on 15 Landsat images
with available binary oil masks

Evaluation metrics (fusion)

Recall: 0.2086
Precision: 0.1253
F1-Score: 0.1566

- 2019-08-26
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

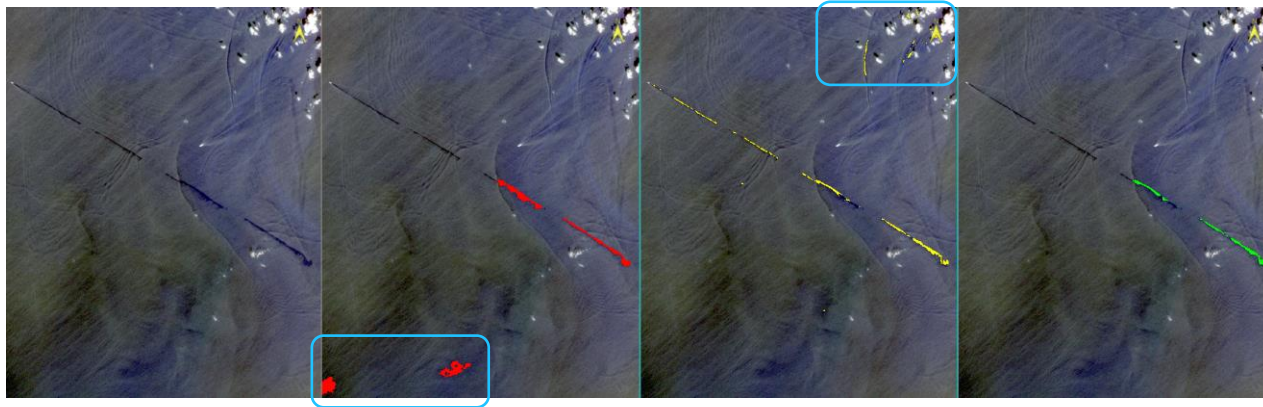
Fusion - Results



-  Thresholding based detections
-  U-Net detections
-  Fusion output (after post-processing)

- 2022-06-28 (top); 2017-06-17 (bottom)
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

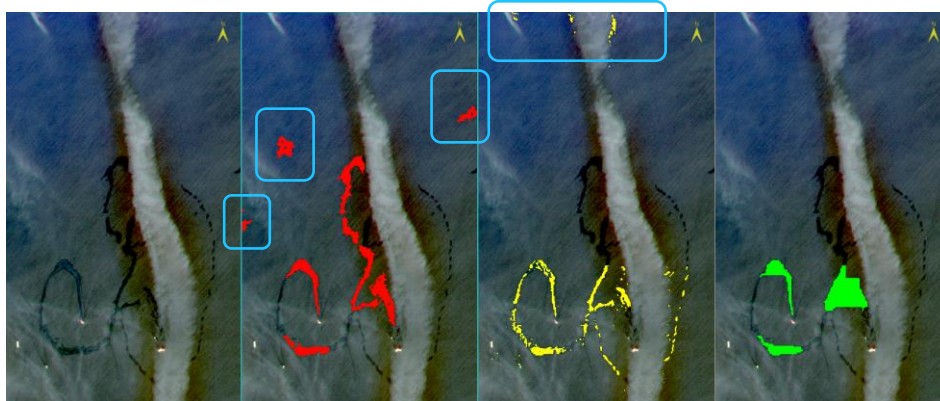
Fusion - Results



False positive reduction

Thresholding: 96.8 %

U-Net: 96.7 %



False positive reduction

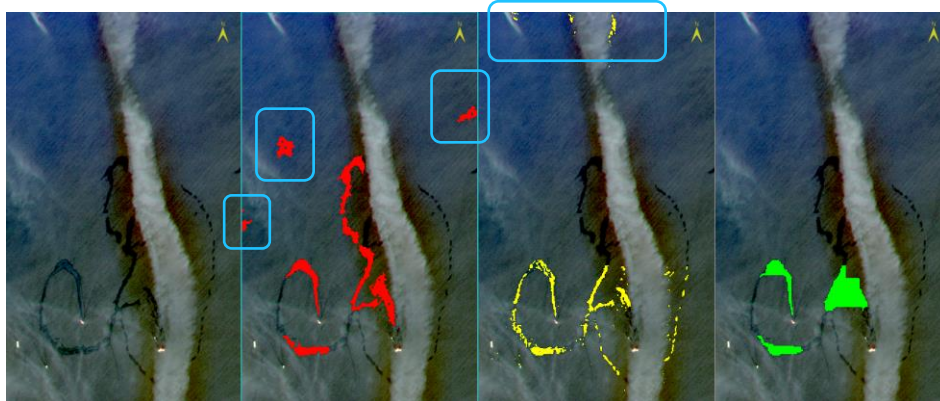
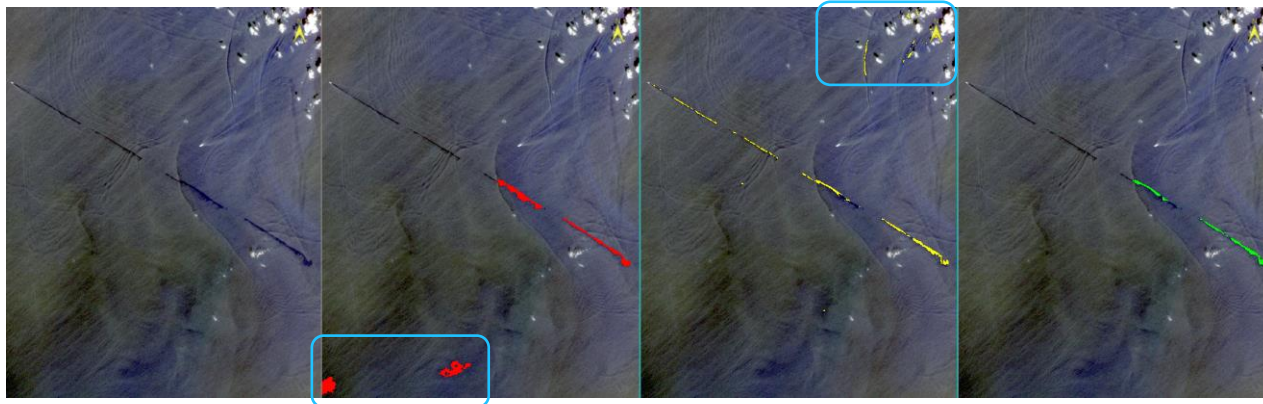
Thresholding: 93.5 %

U-Net: 83.7 %

-  Thresholding based detections
-  U-Net detections
-  Fusion output (after post-processing)

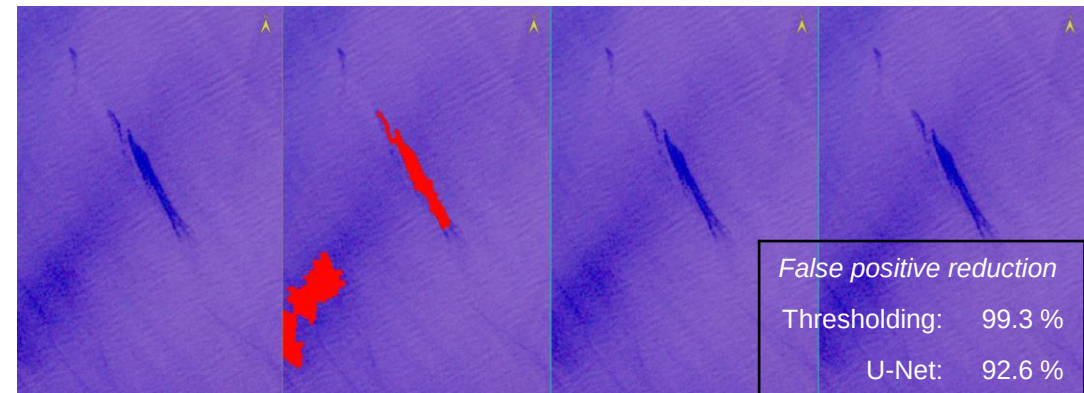
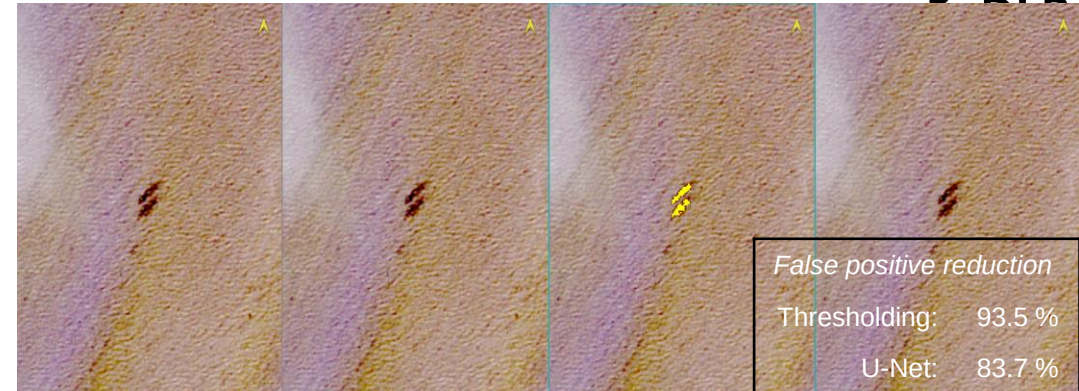
- 2022-06-28 (top); 2017-06-17 (bottom)
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

Fusion - Results



-  Thresholding based detections
-  U-Net detections
-  Fusion output (after post-processing)

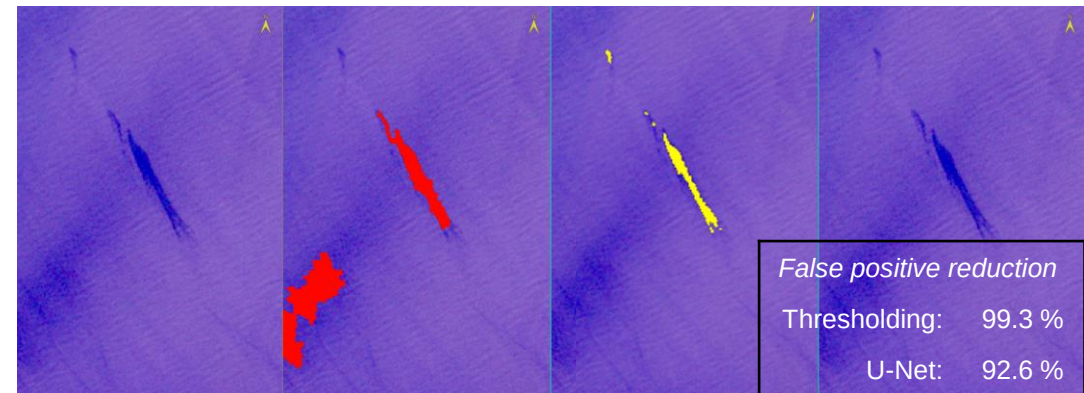
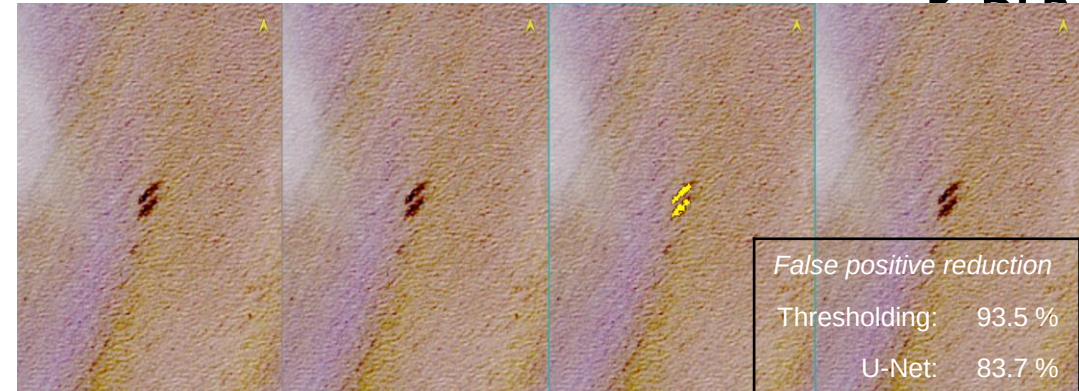
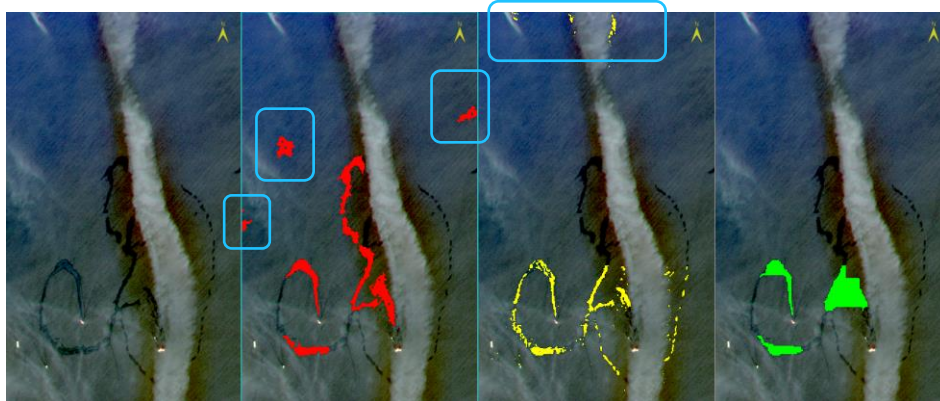
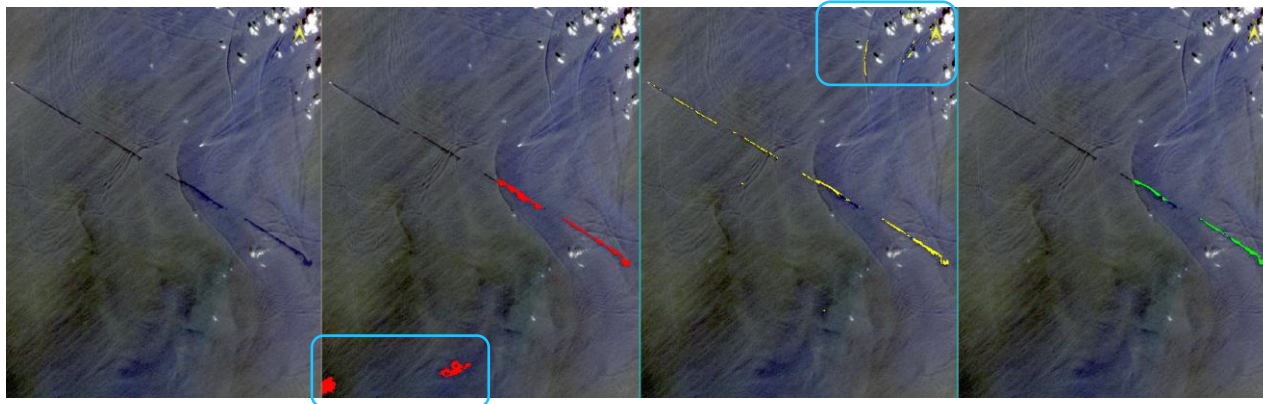
- Left: 2022-06-28 (top); 2017-06-17 (bottom)
- Right: 2024-05-21 (top); 2018-05-07 (bottom)
- False colour images (SWIR2, NIR and blue band), logarithmic stretch



➔ Effective reduction of false positives but at the risk of losing true positives (up to complete loss).

➔ Detection performance is highly scene-dependent!

Fusion - Results



- Thresholding based detections
- U-Net detections
- Fusion output (after post-processing)

- Left: 2022-06-28 (top); 2017-06-17 (bottom)
- Right: 2024-05-21 (top); 2018-05-07 (bottom)
- False colour images (SWIR2, NIR and blue band), logarithmic stretch

➔ Effective reduction of false positives but at the risk of losing true positives (up to complete loss).

➔ Detection performance is highly scene-dependent!

Fusion – Summary & Conclusion

- Reduction of false positive detections, but false positives still remain
- Reduction in size or loss of some oil spills
- Strong scene dependence

Detection challenges

- Low contrast
- High spectral variability due to atmospheric effects (e.g., haze, fog) and cloud related effects
- look-alikes: high spectral similarity to natural phenomena (e.g., **cloud shadows**, wakes, algae, transitions between dark and bright image regions, sun glint)
- Reduced detectability of oil pixels brighter than the surrounding water

Cloud shadow masking

- spectral similarity between oil spills and cloud shadows

Data limitation

- Based on only 92 training patches containing oil pixels (U-Net)
- Based on 48 Landsat scenes (thresholding method)
- Limited generalization

Resolution constraints

- 30 m resolution → small and very narrow oil slicks hard to detect

Thank you!

Olga Schmidt

German Aerospace Center (DLR)
German Remote Sensing Data Center (DFD)
Maritime Safety and Security Lab

Kalkhorstweg 53
17235 Neustrelitz

E-mail: olga.schmidt@dlr.de