



Computer Vision e Deep Learning

A.A. 2023/2024

Relatore:

Prof. Lucia Migliorelli

Supervisori:

Prof. Alessandro Galdelli

Prof. Adriano Mancini

Prof. Stefano Mereu

Studenti:

Angelo Kollcaku

Joshua Sgariglia

Rilevazione di Oil Spill nelle immagini SAR Palsar e Sentinel-1 di bassa qualità



Introduzione

Gli **oil spill** (sversamenti di petrolio) rappresentano una grave minaccia per gli ecosistemi marini, richiedendo un monitoraggio rapido ed efficace per contenere i danni ambientali.

Le immagini **SAR** (Synthetic Aperture Radar) offrono vantaggi significativi:

- Funzionamento indipendente dalle condizioni meteorologiche
- Capacità di acquisizione diurna e notturna
- Elevata sensibilità alle proprietà dielettriche della superficie marina

La sfida principale è rappresentata dal **rumore speckle** e dall'ambiguità visiva che rendono difficile l'identificazione automatica.



Il nostro obiettivo è valutare diverse architetture di deep learning (U-Net e varianti) per la segmentazione semantica di oil spill in immagini SAR a bassa qualità.

Stato dell'arte

Il rilevamento di oil spill rappresenta un task di segmentazione. Al momento presente esistono tre principali insiemi di metodologie utilizzate a tal scopo: algoritmi basati su un threshold, tecniche di machine learning e tecniche di deep learning. I primi due metodi si basano esclusivamente sulla scelta di iperparametri.



1. Algoritmi Basati su Soglia

- Metodi semplici per la segmentazione di aree anomale.
- Approcci rapidi, ma sensibili al rumore e alla variabilità.



2. Machine Learning Tradizionale

- Utilizzo di classificatori come SVM e Random Forest.
- Richiede l'ingegnerizzazione manuale di feature specifiche.



3. Deep Learning

- Modelli avanzati che apprendono automaticamente feature complesse.
- Esempi: UNet++, TransUNet (CNN + Transformer), LightMUNet (State Space Models).

La nostra ricerca si concentra sulla valutazione di UNet, UNet++, TransUNet e LightMUNet su immagini SAR a bassa qualità, introducendo anche varianti leggere come (UNetL, UNet++L) per ottimizzare le risorse computazionali.

Dataset e preprocessing

Dataset

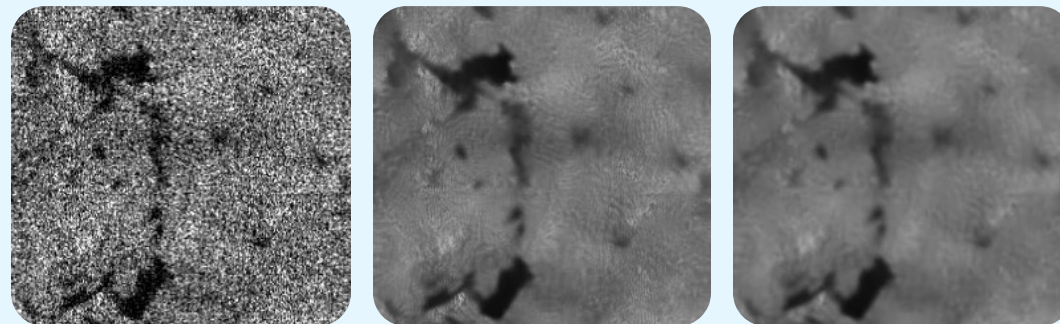
Il dataset SOS corretto nel 2023 raccoglie le immagini scattate da due distinti gruppi di satelliti, Palsar e Sentinel-1A. Si tratta di circa 4000 immagini per ciascun gruppo, già divise in train set e test set.

- Dimensioni: 256×256 pixel
- Forte presenza di rumore speckle
- Maschere binarie per la segmentazione

Pipeline di preprocessing

Abbiamo testato diverse combinazioni di filtri:

- Gaussian
- Median
- Box
- Bilateral
- Fast NL-Means



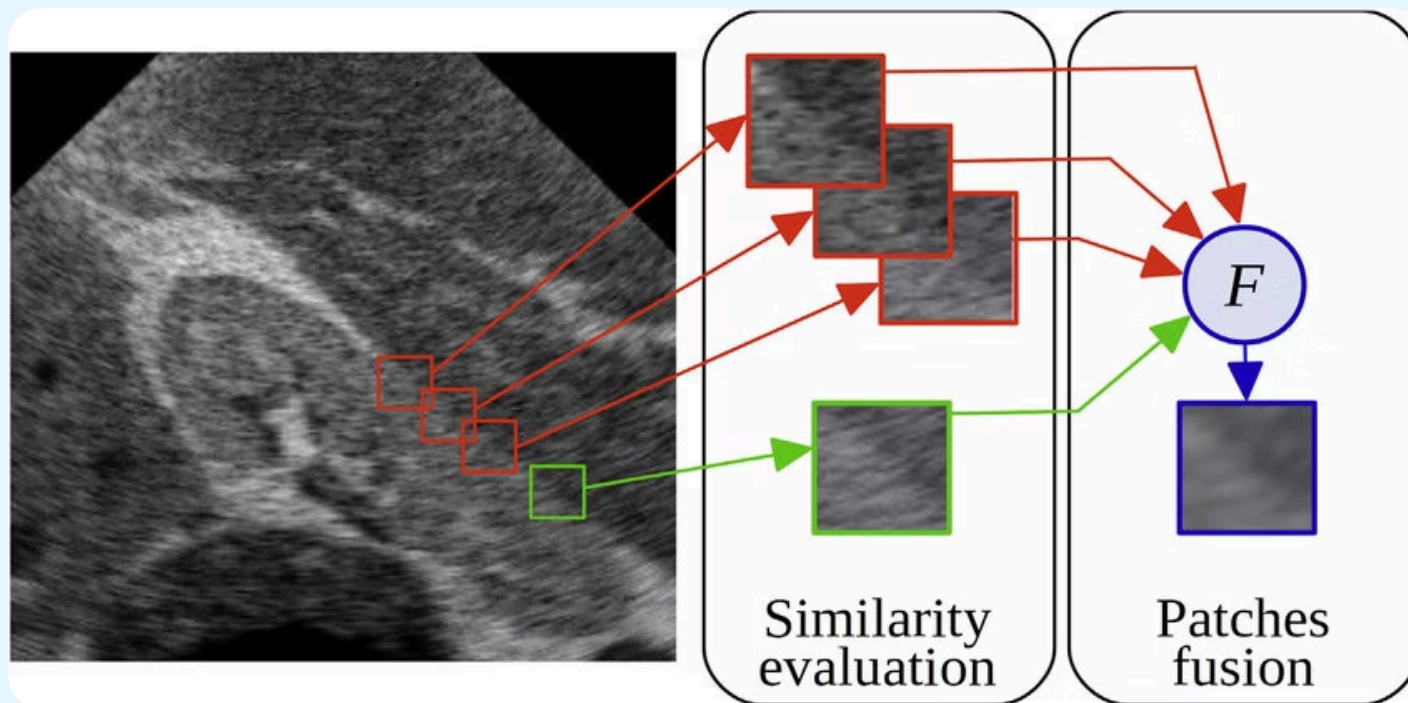
Risultato ottimale

Fast NL-Means seguito da filtro Gaussian

Questa combinazione ha portato a un miglioramento significativo delle metriche:

- mIoU: +1,1%
- mF1-score: +0,7%

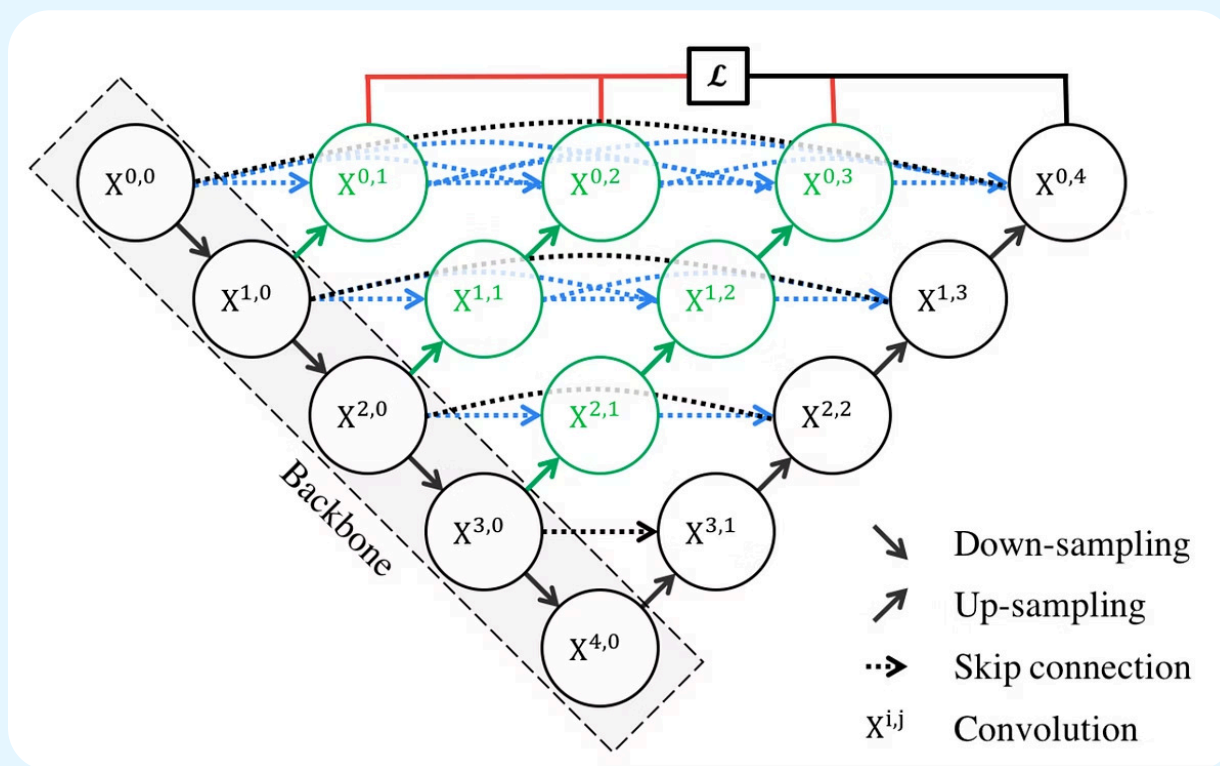
Fast NL-Means



Nel Fast NL-Means, per ogni pixel dell'immagine si considera una patch, cioè un piccolo blocco quadrato centrato su quel pixel. Intorno al pixel si definisce poi una finestra di ricerca più grande, all'interno della quale si cercano altre patch simili. Il confronto tra patch avviene pixel per pixel, calcolando una distanza pesata tra i valori corrispondenti nelle due patch. Le patch più simili ricevono un peso maggiore, e il nuovo valore del pixel viene ottenuto tramite una media pesata dei pixel centrali delle patch presenti nella finestra.

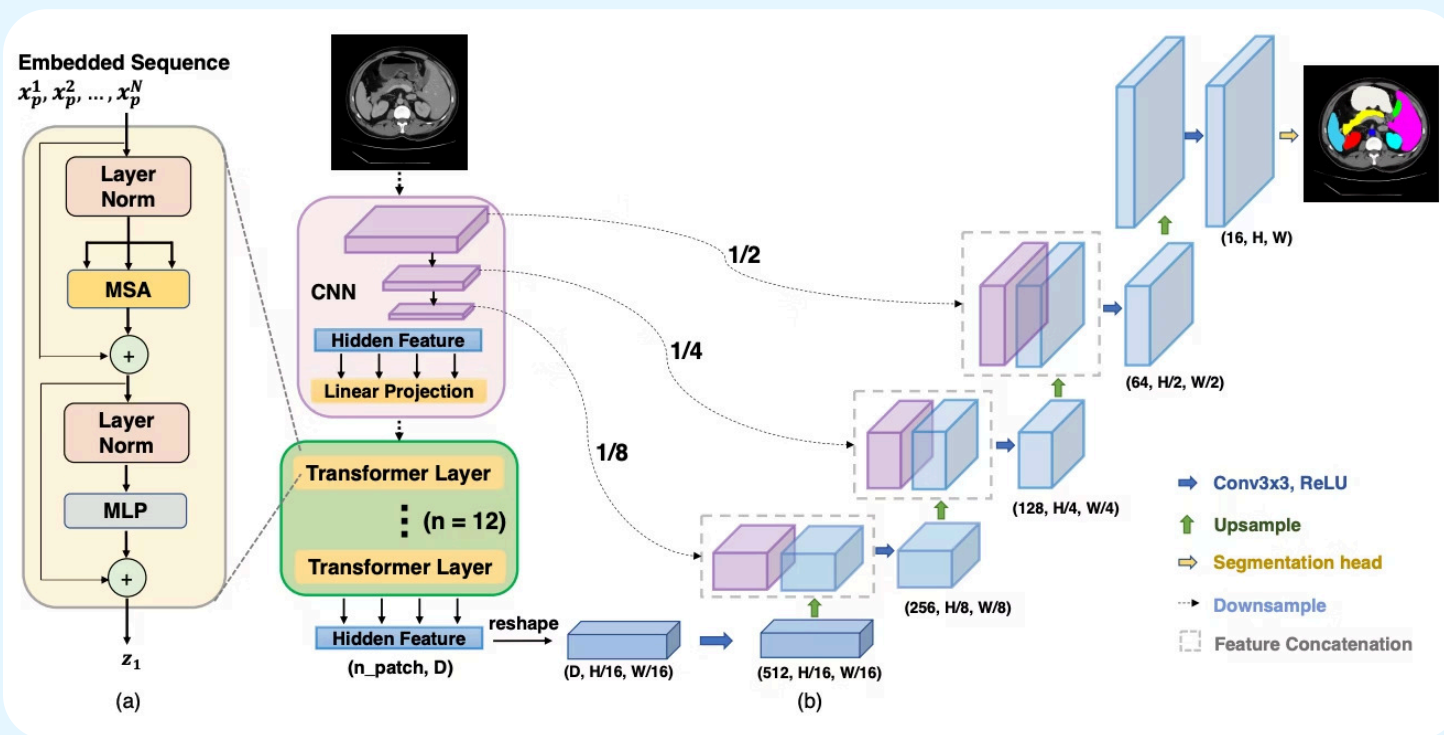
Architetture di rete: struttura e componenti

UNet++



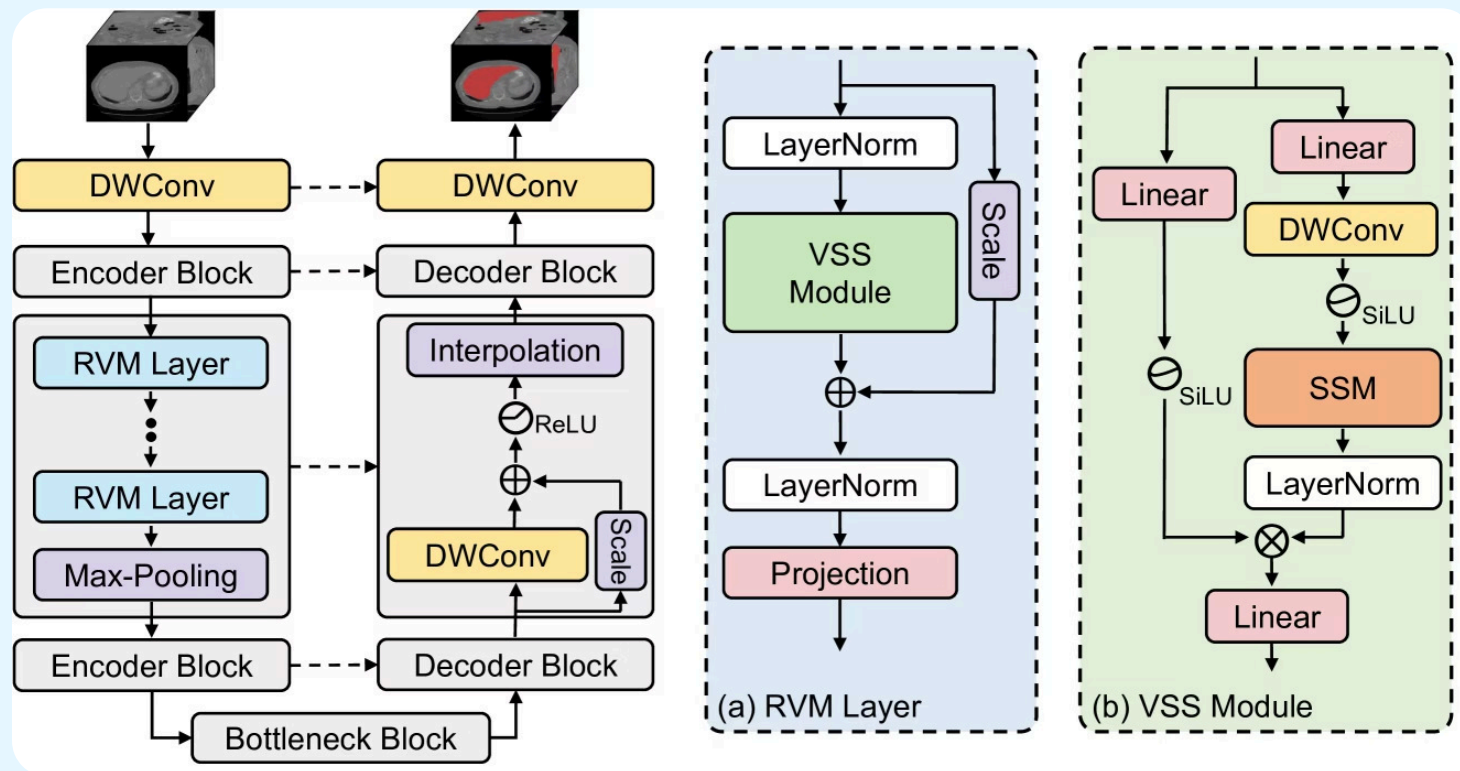
Evoluzione della UNet con connessioni nidificate che formano blocchi densi tra encoder e decoder, migliorando il flusso di informazioni e la capacità di catturare dettagli a diverse scale.

TransUNet



Architettura ibrida che combina CNN per l'estrazione di caratteristiche locali con blocchi Transformer che modellano le dipendenze globali, particolarmente efficace per il riconoscimento contestuale.

LightMUNet



LightMUNet utilizza una struttura a “U”, sostituendo convoluzioni e transformer con blocchi Mamba (SSM) per catturare dipendenze spaziali a lungo raggio. Dopo un'estrazione iniziale con depthwise separable convolution, l'architettura encoder-decoder impiega Residual Vision Mamba Layers e skip connection per combinare profondità semantica e dettagli spaziali.

Training e metriche

Divisione dataset:

80% training

10% validazione

10% test

**Configurazioni
testate: 60**

**Metriche di
valutazione:**

Accuracy

Precision

Recall

F1-score

IoU

| Rete | Epoche | LR | Min LR | Pazienza | Tasso dec | Batch size |
|------------|--------|--------|--------|----------|-----------|------------|
| UNet | 50 | 5e-5 | 1e-6 | 3 - 4 | 0,4 - 0,5 | 8 - 16 -32 |
| UNetL | 50 | 1e-4 | 1e-6 | 3 - 4 | 0,4 - 0,5 | 8 - 16 -32 |
| UNet++ | 40 | 2,5e-5 | 1e-7 | 2 - 3 | 0,4 - 0,5 | 8 - 16 |
| UNet++L | 40 | 5e-5 | 5e-7 | 2 - 3 | 0,4 - 0,5 | 8 - 16 |
| TransUNet | 40 | 1e-4 | 5e-7 | 2 - 3 | 0,3 - 0,5 | 8 - 16 |
| LightMUNet | 40 | 1e-5 | 1e-7 | 2 - 3 | 0,3 - 0,5 | 4 |

Tutte le metriche eccetto Accuracy sono calcolate sia per classe che complessivamente mediando i valori ottenuti per le classi.

Risultati: Palsar

| Rete | Accuracy | mPrecision | mRecall | mF1-score | mIoU |
|----------------------------|----------|------------|---------|-----------|-------|
| UNetL | 93,7% | 89,0% | 88,2% | 88,6% | 80,4% |
| UNet | 93,6% | 88,2% | 89,0% | 88,6% | 80,4% |
| UNet++L | 93,6% | 88,4% | 88,7% | 88,5% | 80,3% |
| UNet++ | 93,7% | 88,5% | 89,0% | 88,8% | 80,6% |
| TransUNet | 93,3% | 87,5% | 89,1% | 88,3% | 79,9% |
| TransUNet preaddestrata | 93,5% | 88,0% | 89,3% | 88,6% | 80,4% |
| LightMUNet | 93,3% | 88,8% | 86,6% | 87,6% | 78,9% |

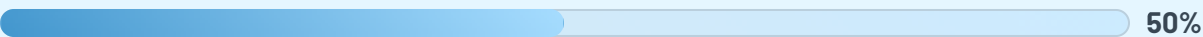
UNet++ ha ottenuto le prestazioni migliori con un miglioramento di mIoU del +0,8% rispetto all'articolo di riferimento, principalmente grazie all'implementazione del preprocessing avanzato.

Risultati: Sentinel

| Rete | Accuracy | mPrecision | mRecall | mF1-score | mIoU |
|----------------------------|--------------|--------------|--------------|--------------|--------------|
| UNetL | 88,7% | 87,9% | 87,1% | 87,5% | 78,0% |
| UNet | 88,8% | 87,7% | 87,6% | 87,7% | 78,3% |
| UNet++L | 88,5% | 87,5% | 87,0% | 87,2% | 77,6% |
| UNet++ | 89,0% | 88,0% | 87,7% | 87,9% | 78,6% |
| TransUNet | 88,7% | 87,9% | 87,0% | 87,4% | 77,9% |
| TransUNet preaddestrata | 89,3% | 88,5% | 88,0% | 88,2% | 79,1% |
| LightMUNet | 87,4% | 86,9% | 85,0% | 85,8% | 75,4% |

TransUNet preaddestrata ha ottenuto le migliori prestazioni con un miglioramento di mIoU del +1,0% rispetto all'articolo di riferimento. Permangono tuttavia difficoltà nella rilevazione di piccole macchie di petrolio, che vengono spesso confuse con il rumore di fondo.

Tempi e impatto ambientale



UNet

Tempo di training: 16,5-16,7 min

Tempo di inferenza: 59-69 ms/immagine



UNet++

Tempo di training: 37,6-40 min

Tempo di inferenza: 67-70 ms/immagine



TransUNet

Tempo di training: 53,8-60,4 min

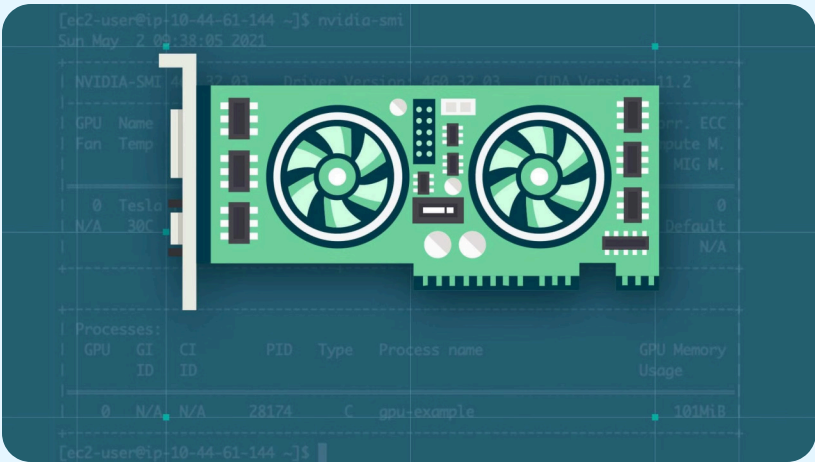
Tempo di inferenza: 105-112 ms/immagine



LightMUNet

Tempo di training: 2,4 ore

Tempo di inferenza: 118-136 ms/immagine



Impatto ambientale

- Emissioni totali: 4,56 kg CO₂
- Correlazione diretta tra complessità del modello e consumo energetico
- Le varianti leggere (UNetL, UNet++L) riducono le emissioni del 15-35% rispetto al originale
- Il training della TransUNet ha emesso il 52-57% di emissioni in meno rispetto alla LightMUNet

È fondamentale considerare anche la sostenibilità ambientale nella scelta dell'architettura ottimale, specialmente per applicazioni che richiedono frequenti ri-addestramenti.

Conclusioni

Risultati chiave

- Miglioramento rispetto allo stato dell'arte:
 - Palsar: +0,8% mIoU (UNet++)
 - Sentinel: +1,0% mIoU (TransUNet)
- Preprocessing fondamentale: Fast NL-Means + Gaussian
- Trade-off precisione/efficienza ben rappresentato dalle varianti leggere

Prospettive future

- Sviluppo di loss personalizzate per migliorare la rilevazione della classe minoritaria
- Esplorazione di modelli Mamba più completi per il processamento di immagini SAR
- Ottimizzazione incentrata sull'efficienza, considerando sia accuratezza che impatto ambientale

Il nostro lavoro dimostra l'efficacia delle architetture moderne di deep learning nella rilevazione di oil spill da immagini SAR a bassa qualità, con particolare enfasi sull'importanza di un preprocessing adeguato.

L'integrazione tra modelli CNN tradizionali e tecnologie emergenti (Transformer, State Space Models) rappresenta la direzione più promettente per i futuri sviluppi in questo campo.

GitHub Repository

