Hybrid Wavelet–CNN Architecture for Rip Current Identification

Yonatan Ben Avraham   Baruch Binyaminov

Introduction

Rip currents are among the most hazardous coastal phenomena worldwide, responsible for the majority of beach-related rescues and a substantial portion of drowning incidents. Despite their global impact, rip currents remain difficult to detect visually because they often lack clear, consistent surface signatures. This challenge is particularly significant on busy public beaches, where timely and accurate identification of dangerous rip channels can directly influence safety operations and prevent loss of life.

Over the past decades, researchers in oceanography, coastal engineering, and computer vision have proposed multiple approaches for rip-current detection. Traditional methods include *in-situ measurements* using ADCPs, GPS drifters, or dye releases, which provide high-accuracy flow information but suffer from limited spatial coverage, high cost, and impracticality for real-time public monitoring. Optical systems based on camera feeds or satellite images offer wider coverage and lower operational costs, yet most existing methods rely heavily on manual interpretation or handcrafted heuristics, making them sensitive to lighting conditions, image quality, and local beach morphology.

Recent progress in *video-based analysis*, *signal processing*, and *deep learning* has motivated increasingly automated strategies. Studies using optical-flow fields, clustering of breaking-wave patterns, or machine-learning classifiers have demonstrated promising results, while deep-learning detectors such as Faster R-CNN, YOLO, or autoencoder-based anomaly models have shown strong performance when large labeled datasets are available. However, deep models suffer from notable drawbacks: substantial data requirements, long training times, sensitivity to environmental variations, and limited interpretability - an important factor in safety-critical applications.

In parallel, classical signal-processing techniques such as the Discrete Wavelet Transform (DWT) have shown strong ability to extract coastal textures, edge structures of breaking waves, and multi-scale patterns associated with rip channels. Wavelets provide localized, multi-resolution analysis and are naturally robust to noise and illumination changes, making them appealing for coastal imagery. Yet, wavelet-based methods alone may struggle with generalization across varied beach environments.

This study proposes a hybrid approach that combines wavelet-based multi-scale feature extraction with a lightweight convolutional neural network and optional attention refinement. The aim is to leverage the interpretability and robustness of wavelet representations together with the discriminative power of modern deep models, while keeping the architecture efficient and suitable for moderate-sized datasets.

Overall, this study highlights the potential of combining classical signal-processing techniques and modern deep-learning components to create accurate, interpretable, and computationally efficient rip-current detectors suitable for real-world coastal monitoring.

Literature Review

Rip currents have been widely studied across coastal science, remote sensing, and computer vision. Early research identified their morphology, circulation patterns, and common visual signatures, forming the basis for later automated detection. Scott et al. [1] provide a comprehensive overview, describing the physical mechanisms that create rip channels and outlining key visual indicators - such as darker gaps in breaking waves or seaward-moving sediment plumes - that guide both human observers and image-based algorithms. An even earlier precursor to automated analysis appears in a 2005 US patent [2], which used handcrafted color and texture cues with rule-based logic or simple neural networks. Although not peer-reviewed, it demonstrates early interest in image-driven detection. The first systematic academic attempts to apply machine learning began with Maryan's 2018 master's thesis and its 2019 peer-reviewed extension [3]. These studies compared classical feature-based models with early convolutional neural networks (CNNs), showing that even shallow CNNs outperform SVM- and Haar-based approaches. They also emphasized the importance of dataset diversity and preprocessing for robust detection. Deep learning methods advanced significantly in subsequent years. De Silva et al. [4] introduced a Faster R CNN framework capable of localizing rip currents in images and video frames, with temporal aggregation improving stability under dynamic surf conditions. Rampal et al. [5] added interpretability by combining MobileNet classifiers with Grad-CAM to highlight the spatial regions driving predictions, providing transparency for operational use. RipViz [6] further incorporated motion dynamics by combining optical flow with an LSTM autoencoder to detect anomalous offshore-directed flow patterns, demonstrating the benefits of modeling water movement over time. In parallel, wavelet-based methods, such as Wang et al. [7], remain relevant for lightweight, computationally efficient detection. Their approach extracts directional water-flow features from images, enhancing breaker fronts while suppressing background noise, making it particularly suitable for resource-constrained or real time systems. Recent work focuses on practical deployment and spatially detailed understanding. RipFinder [8] presents a mobile-ready CNN/YOLO framework for real-time detection on smartphones and low-power devices, while Dumitriu et al. [9] introduced a dense segmentation benchmark using YOLOv8, enabling precise mapping of rip-current contours. Together, these studies trace the evolution from early feature-based methods to deep learning, wavelet approaches, and operational mobile/segmentation systems. This foundation motivates the present work, which aims to build on Wang et al.'s wavelet-based framework, enhancing detection accuracy and robustness while maintaining computational efficiency for practical deployment.

Methodology

1. Dataset

This study utilizes the *Rip Current Monitoring* dataset (Version 4), publicly available on the Roboflow platform. The dataset contains 2,246 coastal images annotated for rip-current presence. The official split is preserved to ensure consistent benchmarking:

- Training: 1,566 images
- Validation: 453 images
- Testing: 227 images

All images are preprocessed by the dataset provider to a uniform resolution of 640×640 pixels, with EXIF orientation removed. Annotations are supplied in YOLOv7 format.

2. Data Preprocessing

To adapt the dataset for binary classification, the following preprocessing pipeline is applied:

1. Label Conversion:
   YOLO annotations are converted into a binary label. Images containing one or more bounding boxes are labeled Rip Current, while images with no annotations are labeled No Rip Current.
2. Normalization:
   Images are kept at 640×640 resolution and normalized to the range [0, 1].
3. Channel Preparation:
   Each sample is prepared in two forms:
   - the original RGB (3-channel) image,
   - a grayscale (1-channel) version used for wavelet computation.
4. Data Augmentation:
   To improve generalization and reduce overfitting, online data augmentation is applied **only to the training set**. The following transformations are used with random probabilities:
   - horizontal flips (to account for left-right symmetry of the surf zone),
   - small rotations (e.g., ±5–10°) to model slight camera tilts,
   - random brightness and contrast jitter to simulate varying illumination,
   - light random cropping or zoom to introduce spatial variability while preserving the surf zone.

   Labels are kept unchanged, and all augmentations are applied consistently to the RGB and grayscale inputs of the same sample.

3. Wavelet-Based Feature Extraction

To incorporate frequency-domain information, a 2-level 2D Discrete Wavelet Transform (DWT) is applied to the grayscale image. Following prior findings in

coastal texture analysis, the Daubechies (db4) wavelet is selected.
The extraction process is defined as follows:

1. Grayscale Input:
   A normalized grayscale tensor is used for the transform.
2. 2-Level Decomposition:
   The DWT produces approximation and detail coefficients at each level. From the second level, we collect:
   - LL_2 (approximation)
   - LH_2 (horizontal detail)
   - HL_2 (vertical detail)
   - HH_2 (diagonal detail)
3. Wavelet Map Construction:
   The four coefficient matrices are resized to 640×640 and stacked into a 4-channel Wavelet Map.

This wavelet representation forms the input to the secondary model branch.

4. Model Architecture

A Dual-Stream Convolutional Neural Network (CNN) is proposed to fuse spatial (RGB) and frequency-domain (wavelet) information.

Stream A: RGB Branch
The RGB image (640×640×3) is processed through a sequence of convolutional layers with ReLU activation and Max Pooling, enabling extraction of spatial and color-based features.

Stream B: Wavelet Branch
The Wavelet Map (640×640×4) is processed by a parallel convolutional stream designed to learn texture- and structure-based cues present in the wavelet coefficients.

Fusion, Optional Attention and Classification
The output feature vectors from both branches are flattened and concatenated. Optionally, a lightweight attention module is applied to this fused feature vector before classification. In this configuration, the concatenated features are passed through a simple attention block (e.g., channel-wise or channel-spatial attention) that reweights feature dimensions according to their estimated importance. The attention-refined representation is then forwarded to the classifier.
Finally, the (optionally attended) fused feature vector is passed through fully connected layers with Dropout regularization. A final Sigmoid unit outputs the predicted probability of rip-current presence.
This setup allows us to compare a **baseline fusion model** (without attention) against an **attention-enhanced variant**, and to quantify the contribution of the attention mechanism in an ablation study.

5. **Training Configuration and Evaluation**

All experiments are conducted using PyTorch with GPU acceleration.

- Loss Function: Binary Cross-Entropy (BCE)
- Optimizer: Adam, initial learning rate 0.001
- Batch Size: 32
- Epochs: Up to 50, with Early Stopping based on validation

loss (patience = 5) Evaluation Metrics

Performance is evaluated on the test set using:

- Accuracy
- Precision
- Recall

Recall is emphasized due to the importance of minimizing false negatives in safety-critical rip-current detection.

References

[1] Scott, T., Masselink, G., Austin, M., & Russell, P. (2009). Rip Current Hazards.

[2] US Patent US20050271266A1 (2005). System and Method for Detecting Rip Currents from Images. [3] Maryan, C. (2018). Detecting Rip Currents from Images (Master's thesis); Maryan, C., Hoque, M., Michael, C., Ioup, E., & Abdelguerfi, M. (2019). Machine Learning Applications in Detecting Rip Channels from Images.

[4] De Silva, A., Mori, I., Dusek, G., Davis, J., & Pang, A. (2021). Automated Rip Current Detection with Region-based Convolutional Neural Networks.

[5] Rampal, N., Shand, T., Wooler, A., & Rautenbach, C. (2022). Interpretable Deep Learning Applied to Rip Current Detection and Localization.

[6] De Silva, A., Zhao, M., Stewart, D., Khan, F., Dusek, G., Davis, J., & Pang, A. (2023). RipViz: Finding Rip Currents by Learning Pathline Behavior.

[7] Wang, H.-M., Doong, D.-J., & Lai, J.-W. (2025). Rip Current Identification in Optical Images Using Wavelet Transform.

[8] Khan, F., De Silva, A., Palinkas, T., Dusek, G., Davis, J., & Pang, A. (2025). RipFinder: Real-Time Rip Current Detection on Mobile Devices.

[9] Dumitriu, M., Tatui, I., Miron, A., Ionescu, R., & Timofte, R. (2025). Rip Current Segmentation: A Novel Benchmark and YOLOv8 Baseline Results.