RipAlert: A Future-Frame-Aware Framework for Rip Current Forecasting and Early Alerting

Meng Wan[1,2*], Qi Su[3,4*], Zhixin Xia[5*], Kanglin Chen[6], Jue Wang[1†], Tiantian Liu[1], Rongqiang Cao[1], Hui Cui[7], Peng Shi[2], Yangang Wang[1], Liqiang Feng[9], Zhenbing Zhao[5]

[1]Computer Network Information Center, Chinese Academy of Sciences
[2]University of Science and Technology Beijing
[3]Peking University
[4]Beijing Academy of Artificial Intelligence
[5]North China Electric Power University
[6]University of California, Davis
[7]China Unicom Software Research Institute
[8]University of Science and Technology Beijing
[9]Laoshan Laboratory
wanmengdamon@cnic.cn, qiisuu@stu.pku.edu.cn, xzx011230@163.com, klichen@ucdavis.edu, wangjue@sccas.cn, liutiantian21a@mails.ucas.ac.cn, caorq@sccas.cn, cuih11@chinaunicom.cn, pshi@ustb.edu.cn, wangyg@sccas.cn, fenglq@qdio.ac.cn, zhaozhenbing@ncepu.edu.cn

Abstract

Rip currents cause over 100 drowning deaths and more than 30,000 rescues annually in the United States, posing a severe threat to beach safety worldwide. However, most existing detection methods are reactive, identifying rip currents only after they form, leaving limited time for intervention. We propose RipAlert, a future-frame-aware framework that forecasts near-future coastal dynamics and proactively identifies rip current risks. We design a region-sensitive optical flow prediction method with a novel entropy-based object detector to capture early-stage reverse-flow anomalies. Unlike static-image approaches, RipAlert leverages temporal motion patterns to detect rip currents up to 5 seconds before they visibly form. To support real-world deployment, we design a lightweight mobile application and release a curated dataset with about 2,000 annotated images. Experiments on the RipVIS benchmark show that our approach achieves state-of-the-art performance. The system has been deployed at high-risk beaches in China, issuing successful early warnings over real-world events. Our work advances AI-driven coastal safety and contributes to SDG 3 (Good Health and Well-Being) and SDG 13 (Climate Action).

Code — https://github.com/AI4SClab/RipAlert
Datasets — https://www.scidb.cn/s/mqIFFb

Introduction

Rip currents are narrow, fast-moving water channels that flow from the shoreline to the open sea and rank among the most dangerous coastal hazards worldwide (Dalrymple et al. 2011; Lee et al. 2016). Often invisible to untrained observers, rip currents account for a substantial portion of surf-related drowning incidents (Castelle et al. 2016). According to national statistics, rip currents cause over 100 drowning deaths annually in the United States (Brewster, Gould, and Brander 2019) and approximately 26 in Australia (Australia 2021), despite hundreds of millions of dollars invested each year in coastal monitoring and rescue operations (Brewster, Gould, and Brander 2019). In certain regions such as the Carolinas, rip-current–related deaths even surpass those caused by hurricanes, tornadoes, and other extreme weather events (Leatherman 2012; National Weather Service 2025). One major reason is that many beachgoers are unaware of the subtle warning signs of rip currents (Shibata 2023). Individuals caught in such currents often panic and attempt to swim directly against the flow, resulting in rapid exhaustion and increased risk of drowning (Cornell et al. 2024). Therefore, Limited safety infrastructure at low-resource beaches highlights the demand for automated early-alerting systems (Castelle et al. 2019).

While the fatality statistics highlight the severity of rip current hazards, the key challenge lies in their dynamic nature and limited early visual indicators (Brander and Scott 2018). Rip currents can develop abruptly and evolve within seconds, leaving insufficient time for reliable human recognition (Shepard, Emery, and La Fond 1941). Traditional safety public education has proven helpful but suffer from limited spatial and temporal coverage, particularly in low-resource settings (Woodward et al. 2015) (Houser et al. 2017). To address the limitations of human-based recognition, early studies proposed detecting rip currents in coastal imagery by extracting handcrafted visual features such as color contrast and texture discontinuities (Pitman et al.

*These authors contributed equally.
†Corresponding author.

2016; Mori et al. 2022). However, almost all existing methods are reactive, as they detect rip currents only after visible cues emerge, limiting the time available for effective warning. Consequently, there is little time to prevent swimmers from entering hazardous areas or to allow lifeguards to intervene preemptively.

Moving from reactive detection to proactive forecasting introduces new technical challenges that existing approaches struggle to address. First, forecasting the near-future sea surface requires models to capture fast-changing wave dynamics within very short time windows (Wantzen, Tharme, and Pypaert 2023). This demands high temporal sensitivity to subtle motion patterns, which many conventional models are not designed to handle (Castelle et al. 2016). Second, it is difficult to detect rip currents before clear visual signs appear, as traditional models rely on static image features (Rampal et al. 2022). Without modeling temporal evolution, these methods often miss early indicators such as flow divergence or directional anomalies (Mori et al. 2022). Third, practical systems must run efficiently on mobile or edge devices, especially in areas lacking reliable infrastructure (Dumitriu et al. 2023) (Wang, Doong, and Lai 2025). This calls for lightweight models with low latency and minimal power consumption to ensure real-time operation in the field. These challenges highlight the need for a new solution that combines short-term video forecasting with real-time detection to enable early alerts.

In this paper, we propose RipAlert, a future-frame-aware framework for real-time rip current forecasting and early alerting, which integrates short-term video prediction with object detection to enable proactive safety interventions along coastlines. The framework is deployed as a lightweight mobile application to support timely alerts in resource-limited beach settings. To the best of our knowledge, this is the first work that combines video-based future frame generation with rip current detection to achieve early warnings in real-world deployment settings. Our key contributions are as follows:

- **Region-sensitive rip current prediction**: We introduce a lightweight video prediction module based on optical flow techniques to forecast the evolution of coastal scenes over the next 3–5 seconds.
- **Entropy-based detector and deployment system**: We design an end-to-end pipeline that couples video prediction with a YOLOv12-based rip current detector and implement it in a mobile-ready application.
- **Open dataset and real-world applicability**: We assemble and annotate a diverse dataset of more than 2,000 rip current images, drawing from both public coastal footage and our own field tests. The annotated dataset is intended for public release to promote reproducible research.

By enhancing rip current awareness and enabling early safety intervention, our research contributes directly to the United Nations Sustainable Development Goals (SDG 3: Good Health and Well-Being, SDG 13: Climate Action).

Related Work

Rip Current Detection in Images

Recent research on rip current recognition has predominantly adopted on conventional object detection and image classification techniques applied to static coastal images. Approaches based on wavelet transform edge detectors, Faster R-CNN and YOLO variants have been extensively studied for identifying dangerous rip channels in coastal imagery (Heric and Zazula 2007; Ren et al. 2016; Khan et al. 2025). For example, YOLO-based architectures have been customized to improve detection accuracy and efficiency through task-specific modifications (Rombado, Orescanin, and Orescanin 2024; Khan et al. 2025). Despite promising results, these methods inherently operate in a reactive manner, relying on current or past frames to identify visible signs of rip currents. Consequently, they suffer from a delayed detection problem, limiting their effectiveness for early warning and preemptive safety interventions.

Temporal Forecasting and Video Prediction

To move beyond delayed detection, some research in video prediction and spatiotemporal modeling have shown potential for anticipating hazards before they become visually prominent. Techniques such as optical flow modeling (Shen, Kerofsky, and Yogamani 2023) and convolutional LSTM (ConvLSTM) architectures (Zheng, Lu, and Zhou 2023) have been employed to predict future frames in domains like autonomous driving and weather forecasting. Similar optical flow methods have also proven effective in short-term motion prediction tasks like cloud movement forecasting (Wan et al. 2025). These models are capable of capturing fine-grained temporal dynamics, which is essential for proactive hazard forecasting. While some efforts have explored real-time visual monitoring pipelines (Dumitriu et al. 2025), existing work has yet to integrate future-frame prediction with real-time rip current detection in a unified framework. To the best of our knowledge, no prior studies explicitly address the fusion of temporal forecasting and active detection for early alerting of rip currents in coastal video feeds.

The Overall Framework

The overall framework of the proposed *RipAlert* is illustrated in Figure 1, comprising four sequential stages: data collection, region-sensitive optical flow prediction, entropy-based detection, and end-to-end application. In the first stage, video streams from drones and coastal cameras are collected at 2 FPS, with contrast enhancement and temporal smoothing applied to suppress noise and improve visual clarity. Next, a dual-frame optical flow algorithm estimates pixel-wise motion, which is segmented into static, turbulent, and reverse-flow regions based on magnitude and direction. These semantic regions guide the adjustment of motion vectors, enabling the synthesis of future frames that emphasize early-stage rip current dynamics. The third stage feeds both historical and semantically enhanced predicted frames into an improved YOLOv12 detector, where we introduce a Content-Aware Entropy Attention module to dynamically adjust spatial focus based on motion complexity. Finally,

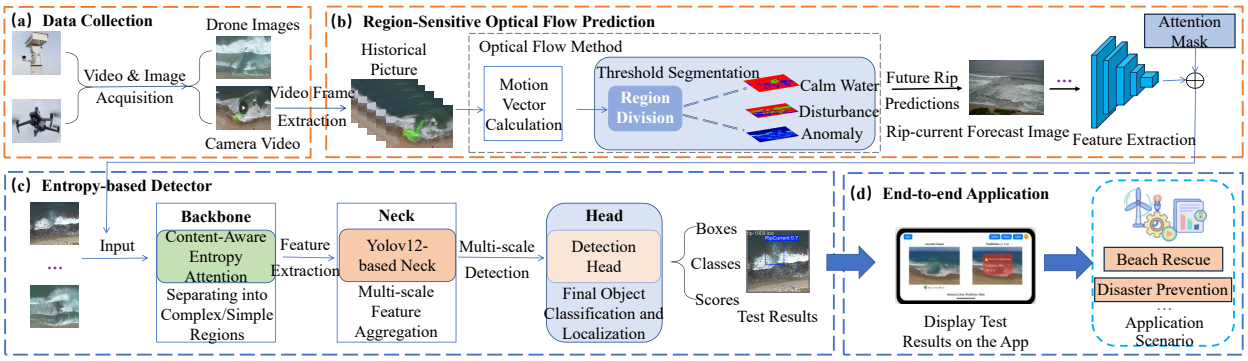


Figure 1: The overall framework of RipAlert.

detection outputs are integrated into a lightweight mobile application that synchronizes predicted risks with user location, providing real-time alerts and supporting beach rescue and disaster prevention. The system forms a cohesive end-to-end pipeline, bridging low-level motion perception and high-level semantic understanding for practical coastal safety applications.

Region-Sensitive Optical Flow Prediction

To enable short-term anticipation of rip current dynamics, we develop a region-sensitive future frame generation method based on dual-frame optical flow. Given two consecutive grayscale frames I_{t-2} and I_{t-1}, we estimate the dense optical flow between them using the Dual TV-L1 algorithm (Zach, Pock, and Bischof 2007; Yang et al. 2024). The flow consists of two components: the horizontal displacement $u(x, y)$ and the vertical displacement $v(x, y)$ at each pixel location (x, y). These are obtained by minimizing an energy function $E(u, v)$ that enforces photometric consistency while penalizing flow discontinuities:

$$E(u,v) = \iint \Big[\big(I_{t-1}(x,y) - I_{t-2}(\tilde{x},\tilde{y})\big)^2 + \lambda\big(|\nabla u(x,y)| + |\nabla v(x,y)|\big) \Big]\, dx\, dy, \tag{1}$$

where $\tilde{x} = x + u(x, y)$ and $\tilde{y} = y + v(x, y)$, and the first term enforces photometric consistency between the warped and reference frames, while the second encourages spatial smoothness of the flow field, weighted by a factor λ.

From the estimated flow components $u(x, y)$ and $v(x, y)$, we further derive the motion magnitude and direction at each pixel:

$$M(x,y) = \sqrt{u^2(x,y) + v^2(x,y)}, \tag{2}$$

$$\theta(x,y) = \arctan 2(v(x,y), u(x,y)), \tag{3}$$

where $M(x, y)$ quantifies motion intensity and $\theta(x, y)$ captures directional changes. These flow descriptors are later used for region segmentation and adaptive refinement.

Region-Aware Motion Segmentation

To enhance the interpretability and robustness of motion modeling, we further segment the flow field into semantically meaningful regions. This is based on two observations: (1) most sea surface areas exhibit stable, low-energy motion; (2) rip currents tend to manifest as reverse or turbulent flow patterns with distinct magnitudes and directions. Therefore, we define three motion types:

- **Static region**: Pixels with low magnitude ($M < \tau_1$), representing calm water.
- **Turbulent region**: Pixels with large magnitude ($M \geq \tau_1$), capturing wave-active zones.
- **Reverse-flow region**: Pixels where the motion direction $\theta(x, y)$ significantly deviates from the dominant orientation $\bar{\theta}$, satisfying $|\theta(x, y) - \bar{\theta}| > \tau_2$ and $M(x, y) > \tau_1$.

where the τ_1 and τ_2 are empirical thresholds for magnitude and directional deviation, and $\bar{\theta}$ denotes the dominant flow direction estimated across the frame. Each pixel is assigned a region label $R(x, y) \in \{0, 1, 2\}$ as follows:

$$R(x,y) = \begin{cases} 0, & \text{if static water} \\ 1, & \text{if turbulent wave} \\ 2, & \text{if reverse flow.} \end{cases} \tag{4}$$

Region-Guided Flow Adjustment

To incorporate semantic priors into motion modeling, we refine the original optical flow field. Let $W(x, y) = (u(x, y), v(x, y))$ denote the original flow vector and $W'(x, y) = (u'(x, y), v'(x, y))$ denote the adjusted flow vector based on the region label $R(x, y)$. The refinement rule is defined as:

$$W'(x,y) = \begin{cases} (0,0), & R(x,y) = 0 \\ W(x,y), & R(x,y) = 1 \\ -\alpha W(x,y), & R(x,y) = 2, \end{cases} \tag{5}$$

where the scalar $\alpha > 1$ acts as an amplification factor to exaggerate reverse motion, improving the model's sensitivity to incipient rip current regions.

Frame Synthesis

Given the refined region-guided flow vector $W'(x, y) = (u'(x, y), v'(x, y))$, we synthesize the next-frame prediction $\hat{I}_t$ by warping the current frame I_{t-1}:

$$\hat{I}_t(x, y) = I_{t-1}(x - u'(x, y),\ y - v'(x, y)), \tag{6}$$

where the $\hat{I}_t(x, y)$ denotes the pixel intensity of the synthesized frame at time t. The warping process incorporates motion semantics to produce future frames that emphasize potential rip current patterns, offering more discriminative supervision for downstream detection.

Entropy-Based Detector

Modern one-stage detectors such as YOLOv12 (Tian, Ye, and Doermann 2025) adopt lightweight attention mechanisms to expand the receptive field efficiently. However, these methods apply identical operations across all spatial locations, regardless of local content complexity. To better handle the sparse yet highly textured nature of rip current regions embedded in smooth coastal backgrounds, we propose a Content-Aware Entropy Attention (CEA) module. By computing entropy over local windows, CEA dynamically routes high-entropy regions to deformable attention and low-entropy regions to depth-wise convolutions. This selective processing enhances foreground discrimination while maintaining comparable computational cost to standard area attention.

Content-Aware Entropy Attention (CEA)

The Content-Aware Entropy Attention module selectively applies deformable attention to high-entropy regions associated with turbulent rip flows, while assigning lightweight depth-wise convolutions to low-entropy areas such as smooth water or sand. By dynamically routing computation based on local entropy, it ensures efficient processing and sharper focus on visually complex rip current patterns.

Pre-processing. In the first step of the CEA module, the input feature map $\mathbf{F} \in R^{C\times H\times W}$, where C is the number of channels, H is the height, and W is the width, is processed using a per-pixel convolution to project it to a working dimension. Positional encodings $\mathbf{P}$ are added to $\mathbf{F}$ to incorporate spatial awareness. The resulting feature map, denoted $\mathbf{F}'$, is then used to predict an offset map $\mathbf{F}_{\text{offset}}$ which captures the displacement needed for deformable attention, allowing the model to focus on intricate areas such as rip currents. Additionally, lightweight spatial attention $\mathbf{SA}$ and channel attention $\mathbf{CA}$ are computed from $\mathbf{F}'$. These are used in later stages to refine the features of simpler regions while leaving complex regions to deformable attention.

Entropy-based Window Classification. After the preprocessing step, the feature map $\mathbf{F}'$ is partitioned into non-overlapping windows of size $M \times M$. Each window is reshaped to form $\mathbf{F}_v \in R^{N_w\times M^2\times C}$, where $N_w = \frac{HW}{M^2}$ represents the total number of windows. For each window, a softmax operation is applied to normalize the token distribution, and the Shannon entropy $\mathbf{E}_n$ is computed for each window:

$$\mathbf{E}_n = -\sum_{i=1}^{M^2} P_n(i) \log_2 P_n(i), \tag{7}$$

where $P_n(i)$ is the normalized probability of the i-th token in the n-th window, and $\mathbf{E}_n$ represents the entropy of the window. Each window's entropy value is then compared with the mean entropy across all windows to determine its complexity level. Windows whose entropy lies between one-half and the full value of the mean are categorized as complex regions (e.g., rip currents or waves with strong textures), while those with entropy below half of the mean are considered simple regions (e.g., smooth water or sand). Corresponding binary masks, denoted as $\mathbf{Mask}_{\text{complex}}$ and $\mathbf{Mask}_{\text{simple}}$, are subsequently generated to enable separate processing of complex and simple regions in the following stages.

Dual-branch Token Mixing

Complex Region Feature Enhancement. Once the complex and simple regions are classified, the complex regions undergo deformable window attention to capture detailed, object-centric features. Deformable attention allows the model to focus on relevant regions in the complex areas by shifting the sampling grid according to the offset map $\mathbf{F}_{\text{offset}}$, which was computed during pre-processing. The resulting feature map $\mathbf{X}$ is reshaped and used to derive query ($\mathbf{Q}$) and key ($\mathbf{K}$) embeddings, which are then used to compute self-attention over the complex region values.

$$\mathbf{V}_{\text{complex}} = \mathbf{Mask}_{\text{complex}} \odot \mathbf{F}, \tag{8}$$

$$\mathbf{V}_{\text{complex}} \leftarrow \text{Softmax}\left(\frac{\mathbf{QK}^\top}{\sqrt{d}}\right)\mathbf{V}_{\text{complex}}. \tag{9}$$

In the case of rip currents (complex regions), the deformable attention mechanism enables the model to adaptively attend to relevant parts of the image, enhancing the features in complex regions where rip currents appear.

Simple Region Feature Refinement. For the simple regions, a simpler operation such as depth-wise convolution is applied. The simple region features $\mathbf{V}_{\text{simple}}$ are multiplied by the simple region mask $\mathbf{Mask}_{\text{simple}}$ to select only the simple region tokens. These features are then processed through spatial attention $\mathbf{SA}$ to further refine them before they are merged with the complex region features.

$$\mathbf{V}_{\text{simple}} = \mathbf{Mask}_{\text{simple}} \odot \mathbf{F}, \tag{10}$$

$$\mathbf{V}_{\text{simple}} \leftarrow \mathbf{SA} \odot \text{DWConv}_{3\times3}(\mathbf{V}_{\text{simple}}). \tag{11}$$

Feature Fusion. Finally, the enhanced complex region features $\mathbf{V}_{\text{complex}}$ and the refined simple region features $\mathbf{V}_{\text{simple}}$ are combined to produce a unified output feature map $\mathbf{V}_{\text{out}}$. This fusion process ensures that both complex and simple region information are integrated into a single representation. A final channel attention $\mathbf{CA}$ is applied to emphasize discriminative features and further refine the output:

$$\mathbf{V}_{\text{out}} = (\mathbf{V}_{\text{complex}} + \mathbf{V}_{\text{simple}}) \odot \mathbf{CA}. \tag{12}$$

The resulting $\mathbf{V}_{\text{out}}$ is passed to the next layer of the network, where it will be used for the final detection or segmentation

tasks. This unified feature map includes both complex, detailed features from the complex rip currents and the simple background information, making it suitable for accurate object detection in the context of rip currents.

Experimental Evaluations

Experimental Settings

Dataset. We adopt two complementary datasets for training and evaluation. (1) The *Rip Current Segmentation Benchmark* (Dumitriu et al. 2025) provides 2,466 annotated rip-current images and 1,307 negatives, along with 17 test videos (24,295 frames, 30 FPS) in resolutions of 3840×2160 and 1920×1080, covering diverse coastal scenes. (2) We further construct a custom dataset consisting of 2,143 annotated frames, collected from both publicly available coastal imagery and our own drone and shore surveillance in high-risk areas. The dataset focuses on early-stage rip current patterns and is annotated with bounding boxes and confidence levels. It will be publicly released to support reproducible research.

Training Configuration. All models are implemented in PyTorch and trained using the Adam optimizer with an initial learning rate of 1×10^{-4}. Each video sequence is preprocessed to a fixed resolution of 512×512, and a random horizontal flip is applied during training for augmentation. The batch size is set to 8, and training is conducted on a single NVIDIA A100 GPU for 100 epochs. The optical flow and future frame generation modules are pre-trained separately before joint fine-tuning with the detection head.

Baselines. We select five competitive models as baselines, including TV-L1 (Zach, Pock, and Bischof 2007; Yang et al. 2024), Lucas-Kanade (Plyer, Le Besnerais, and Champagnat 2016), YOLOv8 (Sohan, Sai Ram, and Rami Reddy 2024), RT-DETR (Kong, Shang, and Jia 2024), YOLOv12 (Tian, Ye, and Doermann 2025).

Parameter Details. We divide the dataset into a training set and a test set in an 8:2 ratio. The video frame rate is downsampled to 2 FPS for efficiency, and each sample includes a 5-frame historical sequence for future frame prediction. The optical flow prediction module is trained for 30 epochs using the Adam optimizer with an initial learning rate of 0.001 and a cosine annealing schedule. The input frames are resized to 512×512, and the output future frame corresponds to 2.5 seconds ahead. In the motion decomposition module, we compute the optical flow field and segment it into static, turbulent, and reverse-flow regions using magnitude-direction clustering. A region-wise motion attention map $\mathcal{A}$ is generated to guide downstream detection. For the YOLOv12-based detection module, the input is a three-channel composite $\mathbf{X}_t = \text{Concat}(I_{t-1}, \hat{I}_t, \mathcal{A})$. The network is trained for 50 epochs with a batch size of 16, using a one-cycle learning rate schedule starting at 1×10^{-4}. The object confidence threshold is set to 0.5 and the NMS IoU threshold is set to 0.6.

Evaluation Metrics

Peak Signal-to-Noise Ratio (PSNR). PSNR measures the pixel-level error between the predicted and actual images, and is defined as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\text{MSE}} \right), \tag{13}$$

where MAX_I is the maximum possible pixel value of the image, and MSE is the mean squared error between the predicted frame and the ground-truth frame. A higher PSNR value indicates smaller pixel-wise errors and better image quality.

Structural Similarity Index (SSIM). SSIM reflects the perceived visual quality by modeling the human visual system, measuring the similarity in luminance, contrast, and structural information. It is defined as:

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \tag{14}$$

where x and y denote the predicted and ground-truth frames, respectively. When $\alpha = \beta = \gamma = 1$, SSIM ranges from 0 to 1, with higher values indicating stronger structural consistency and visual fidelity.

Intersection over Union (IoU). The Intersection over Union (IoU) measures the similarity between two bounding boxes, defined as the ratio of the intersection area to the union area:

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{15}$$

where A and B are the predicted and ground truth bounding boxes, respectively. This metric is used to calculate mAP at different IoU thresholds.

Average Precision (AP). The AP for a single class is calculated by ranking predictions by confidence and computing the area under the precision-recall curve. The formula for AP is:

$$\text{AP} = \sum_n (\text{Recall}_n - \text{Recall}_{n-1}) \cdot \text{Precision}_n, \tag{16}$$

where n corresponds to the recall levels at which precision is calculated.

Mean Average Precision (mAP). The mAP is the average of the AP values at different IoU thresholds. For mAP_{50}, AP is computed at an IoU threshold of 0.5. For $\text{mAP}_{50:95}$, AP is averaged across IoU thresholds from 0.5 to 0.95 in increments of 0.05:

$$\text{mAP}_{50} = \text{AP}_{IoU=0.5}, \tag{17}$$

$$\text{mAP}_{50:95} = \frac{1}{k} \sum_{i=1}^{k} \text{AP}_{IoU=i} \quad \text{where} \quad k = 10, \tag{18}$$

with k being the number of IoU thresholds (10 in this case, corresponding to 0.5, 0.55, ..., 0.95).

Region-Sensitive Prediction Results

To evaluate the quality of predicted frames generated by different optical flow algorithms (Region-Sensitive Methods (Ours), TV-L1, and Lucas-Kanade), we adopt PSNR and SSIM to quantify the difference between predicted frames and ground-truth frames from both pixel-wise and perceptual perspectives.

Method	PSNR	SSIM
TV-L1	30.6631	0.7226
Lucas-Kanade	30.6226	0.7168
Region-Sensitive Methods (Ours)	30.8708	0.7539

Table 1. Quantitative evaluation of predicted frame quality.

As shown in Table 1, Our methods achieve a PSNR of 30.8708 and an SSIM of 0.7539. The higher PSNR suggests that the predicted frames by the method have less noise and distortion compared to the ground-truth frames. The elevated SSIM indicates that the structural information of the predicted frames, especially in complex texture regions like disturbed water surfaces, is more consistent with the real scenes. Overall, the integration of region segmentation into the TV-L1 algorithm effectively enhances the accuracy and robustness of optical flow estimation, making it more suitable for scenarios with complex motion and texture changes, such as wave-related image sequences.

Optical Flow Prediction Visualization. As shown in Figure 2, using optical flow to predict the movement of rip currents. The large image on the far left shows the original image from the last frame of the video. The green arrows overlaid on the actual image indicate the estimated optical flow toward the next time step. The direction and length of the arrows represent the direction and magnitude of the movement, respectively. The predicted rip current image is displayed below the observation results for comparison. The delta image on the far right shows the pixel differences between the predicted and actual images 3 seconds later.

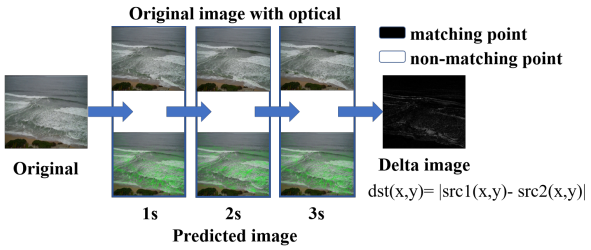


Figure 2: Optical Flow Prediction Visualization.

Entropy Based Detector Results

As shown in Table 2, our enhanced detector achieves state-of-the-art performance across all evaluated metrics. The baseline YOLOv12 model reaches a precision of **90.7%** and mAP_{50} of **90.96%**, already outperforming YOLOv8 variants and RT-DETRv2 on this challenging coastal safety dataset. Our full model, which integrates the proposed CEA module, further boosts performance to a precision of **93.36%**, a recall of **88.44%**, and an mAP_{50} of **94.68%**. This performance gain highlights the effectiveness of allocating computational attention based on local visual complexity. In particular, CEA enables the model to focus on sparse but highly textured regions, such as turbulent surf zones and incipient rip currents. At the same time, CEA reduces noise from visually redundant backgrounds like sand and open water. These improvements are especially valuable in rip current detection, where early-stage features are often subtle and spatially localized. Moreover, the noticeable gap between mAP_{50} and $mAP_{50:95}$ (approximately 46%) across all methods indicates that high-IoU object localization remains a challenge. While our model achieves the best results, this observation suggests room for future work on bounding-box refinement, possibly through instance segmentation or uncertainty-aware detection strategies.

Model	Precision	Recall	mAP_{50}	$mAP_{50:95}$	Fitness
YOLOv8n-seg	0.8827	0.8467	0.8894	0.4575	0.5209
YOLOv8s-seg	0.8762	0.8467	0.8925	0.4666	0.5195
YOLOv8m-seg	0.8804	0.8504	0.8992	0.4748	0.5223
YOLOv8l-seg	0.8904	0.8497	0.8972	0.4685	0.5190
YOLOv8x-seg	0.8874	0.8426	0.9019	0.4745	0.5145
RT-DETRv2	0.8424	0.7590	0.8770	0.4290	0.4738
YOLO12	0.9067	0.8521	0.9096	0.4788	0.5219
Ours	0.9336	0.8844	0.9468	0.4849	0.5311

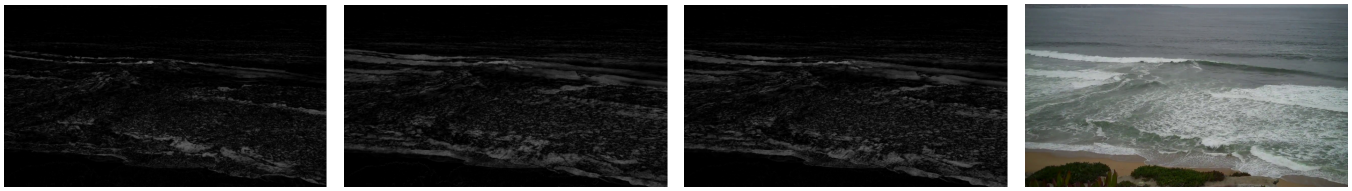
Table 2. Performance of the baselines results on the datasets.

Trade-off Analysis of Forecast Horizon

To determine the optimal forecast horizon for a reliable early warning system, we analyze the trade-off between prediction fidelity, system latency, and the resulting temporal lead. As presented in Table 3, a 2-second forecast horizon establishes a strong baseline, which provides a 1.839-second temporal lead from a total latency of only 161ms (with 12ms transmission time, 128ms forecast time and 21ms detect time). The value of 0.75 SSIM refers to high prediction fidelity. While extending the forecast to 4 seconds yields a substantial 3.706s lead, this comes at the cost of fidelity, which degrades to 0.66 SSIM. This degradation becomes more pronounced at the 6-second horizon, where a precipitous drop in SSIM to 0.56 may render the warning unreliable, despite the extensive 5.573s lead. Consequently, our analysis indicates that a 2-4 second horizon provides a balance of foresight and accuracy for effective early warnings.

Ablation Study

Effect of Region-Aware Segmentation in Optical Flow. We evaluate the benefit of incorporating region-aware segmentation into optical flow-based frame prediction. As shown in Figure 3, we compare our Region-Sensitive TV-L1 variant with two conventional methods: standard TV-L1 and Lucas-Kanade. Difference maps between predicted and



(a) TV-L1+Region (b) TV-L1 (c) Lucas-Kanade (d) Ground Truth

Figure 3: Visualization of predicted frames by different optical flow methods and the corresponding ground truth.

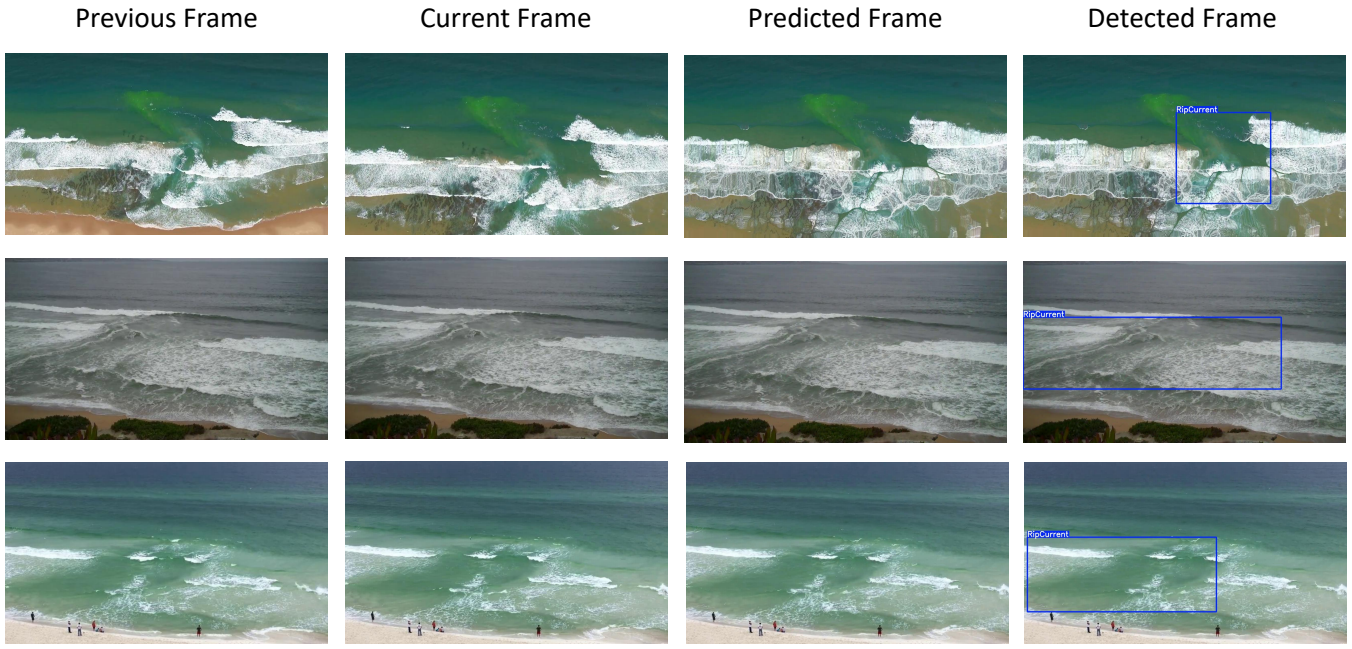


Figure 4: Illustration of the deployed rip current forecasting system interface in a high-risk beach zone. The figure shows the sequence of frames involved in the model's prediction and detection process. From left to right: (1) **Previous Frame**: The initial frame used as input to the model, (2) **Current Frame**: The subsequent frame processed alongside the previous frame to predict the next, (3) **Predicted Frame**: The frame predicted by the model using optical flow techniques, and (4) **Detected Frame**: The predicted frame with detected features highlighted, demonstrating the model's detection capabilities.

Horiz.	Trans.	Frcst.	Detect.	T. Lead	SSIM
2	12	128	21	1.839	0.75
4	13	260	21	3.706	0.66
6	12	394	21	5.573	0.56

Table 3. Trade-off Analysis of Forecast Horizon. **Column definitions:** Horiz. (Horizon, s), Trans. (Transmission Time, ms), Frcst. (Forecast Time, ms), Detect. (Detection Time, ms), and T. Lead (Temporal Lead, s). All time metrics are reported in ms or s as indicated.

ground-truth frames are visualized to highlight spatial error distributions. Compared to standard TV-L1 and Lucas-Kanade methods, our region-sensitive variant produces noticeably smaller prediction errors with fewer intense artifacts, especially in dynamic regions near the wave-breaking boundaries.

Effect of Entropy-Based Attention in Detection. To isolate the contribution of the CEA, we compare the original YOLOv12 model with our models. As reported in Table 2, the inclusion of CEA improves the mAP_{50} from 90.96% to 94.68%, with noticeable gains in both precision and recall. These results demonstrate that CEA effectively enhances detection sensitivity in complex marine environments by prioritizing high-entropy regions and suppressing noise in smooth backgrounds.

Robustness to Illumination and Weather Variations. To further assess the robustness of our proposed model, RipAlert, we conducted an extensive evaluation on a supplementary dataset comprising 1,200 frames captured under challenging environmental conditions: cloudy (500 frames), low-light (400 frames), and glare (300 frames). The evaluation focused on two key tasks: future-frame prediction and object detection. As shown in Table 4, SSIM slightly decreases under low-light but overall remains stable. From

Table 5, mAP shows larger variation, mainly decreasing in low-light due to reduced visual contrast, yet our model still consistently outperforms YOLOv12 and RT-DETR across all illumination settings.

Method	Clear	Cloudy	Low-light	Glare
TV-L1	0.72	0.68	0.62	0.67
Ours	0.75	0.72	0.67	0.71

Table 4. Future-frame prediction performance (SSIM).

Method	Clear	Cloudy	Low-light	Glare
RT-DETRv2	0.874	0.772	0.574	0.811
YOLOv12	0.901	0.803	0.582	0.845
Ours	0.947	0.832	0.645	0.904

Table 5. Detection performance comparison (mAP_{50}).

Qualitative Results and System Visualization

To complement the quantitative metrics reported earlier, we present a visual analysis of the model's inference process in Figure 4. The image sequence provides a qualitative example of how the predicted future frame contributes to early and accurate rip current detection. As shown, the predicted frame captures subtle motion patterns that are not yet apparent in the current input, allowing the detector to highlight risk regions ahead of time. The example verifies the effectiveness of our region-sensitive optical flow module and entropy-based detection head from a visual standpoint. Compared to static image baselines, the model identifies rip current signatures before they are visibly formed, supporting the quantitative improvements in both SSIM and mAP.

Deployment and Social Impact

Rip Current Forecasting System Interface

To support real-time beach safety management, we developed a future-frame-aware rip current forecasting system, jointly piloted with the Institute of Oceanology, Chinese Academy of Sciences (IOCAS). As shown in Figure 5, the platform integrates video-based motion prediction, early warning alerts, and mobile app interfaces. It has been deployed at selected high-risk coastal zones in Shandong and Fujian Provinces, enabling preemptive detection of rip current events 5–10 seconds in advance. The system provides both on-site visual alerts and mobile notifications to beachgoers and coastal supervisors. As a field-deployable, lightweight solution, the system demonstrates how AI-enabled sensing can strengthen public safety infrastructure and support intelligent coastal governance.

Real-World Social Impact

The proposed future-aware rip current forecasting framework has been piloted in collaboration with the IOCAS in high-risk coastal zones of Shandong. During an initial 3-month pilot deployment at Qingdao beach, the system successfully issued early alerts for 32 distinct rip current events, reducing potential drowning risks through real-time push notifications and on-site display screens. A companion mobile app has been developed to serve coastal supervisors, supporting timely alerts, visual overlays, and event logging on low-power edge devices. To evaluate its practical effectiveness, a post-deployment survey was conducted with 20 coastal safety supervisors. Field feedback showed a 90% satisfaction rate. Users specifically reported high satisfaction with the system's alert timing, user interface clarity, reliability, the actionability of the information, and its overall utility. By enabling affordable, deployable early warning in resource-limited coastal regions, the framework contributes to SDG 3 and SDG 13, and supports LNOB principles.

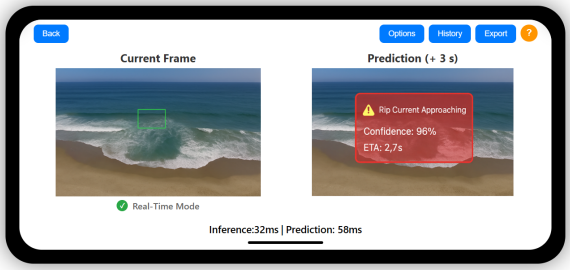


Figure 5: Rip current Alerting Application interface.

Conclusion

This paper introduces RipAlert, a future-aware rip current forecasting framework that combines optical flow-based frame prediction and motion-guided detection. By identifying early-stage reverse-flow patterns, RipAlert enables preemptive alerts ahead of time. Experiments show its superior performance over frame-wise baselines. Future work will focus on broader deployment and integration with real-time coastal monitoring systems.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was supported by National Key R&D Program of China (No.2025YFE0102600). We thank VenusAI Platform (http://data.aicnic.cn) for kindly providing the GPU clusters for model training.

References

Australia, S. L. S. 2021. National Coastal Safety Report. https://www.marinebusinessnews.com.au/wp-content/uploads/2021/09/Surf-Life-Saving-2021-Report.pdf.pdf. Accessed: 2025-12-04.

Brander, R.; and Scott, T. 2018. Science of the rip current hazard. In *The science of beach lifeguarding*, 67–85. CRC Press.

Brewster, B. C.; Gould, R. E.; and Brander, R. W. 2019. Estimations of rip current rescues and drowning in the United States. *Natural Hazards and Earth System Sciences*, 19(2): 389–397.

Castelle, B.; Scott, T.; Brander, R.; McCarroll, J.; Robinet, A.; Tellier, E.; De Korte, E.; Simonnet, B.; and Salmi, L.-R. 2019. Environmental controls on surf zone injuries on high-energy beaches. *Natural Hazards and Earth System Sciences*, 19(10): 2183–2205.

Castelle, B.; Scott, T.; Brander, R.; and McCarroll, R. 2016. Rip current types, circulation and hazard. *Earth-Science Reviews*, 163: 1–21.

Cornell, S.; Brander, R. W.; Roberts, A.; Koon, W.; Peden, A. E.; and Lawes, J. C. 2024. 'I actually thought that I was going to die': Lessons on the rip current hazard from survivor experiences. *Health promotion journal of Australia*, 35(2): 551–564.

Dalrymple, R. A.; MacMahan, J. H.; Reniers, A. J.; and Nelko, V. 2011. Rip currents. *Annual Review of Fluid Mechanics*, 43(1): 551–581.

Dumitriu, A.; Tatui, F.; Miron, F.; Ionescu, R. T.; and Timofte, R. 2023. Rip current segmentation: A novel benchmark and yolov8 baseline results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1261–1271.

Dumitriu, A.; Tatui, F.; Miron, F.; Ralhan, A.; Ionescu, R. T.; and Timofte, R. 2025. RipVIS: Rip Currents Video Instance Segmentation Benchmark for Beach Monitoring and Safety. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 3427–3437.

Heric, D.; and Zazula, D. 2007. Combined edge detection using wavelet transform and signal registration. *Image and Vision computing*, 25(5): 652–662.

Houser, C.; Trimble, S.; Brander, R.; Brewster, B. C.; Dusek, G.; Jones, D.; and Kuhn, J. 2017. Public perceptions of a rip current hazard education program:"Break the Grip of the Rip!". *Natural hazards and earth system sciences*, 17(7): 1003–1024.

Khan, F.; Stewart, D.; de Silva, A.; Palinkas, A.; Dusek, G.; Davis, J.; and Pang, A. 2025. RipScout: Realtime ML-Assisted Rip Current Detection and Automated Data Collection Using UAVs. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Kong, Y.; Shang, X.; and Jia, S. 2024. Drone-DETR: Efficient small object detection for remote sensing image using enhanced RT-DETR model. *Sensors*, 24(17): 5496.

Leatherman, S. P. 2012. Rip currents. In *Coastal Hazards*, 811–831. Springer.

Lee, G.; Hong, S.; Lee, C.; Kim, J.; and Lee, J. 2016. Rip current zoning map to manage safety at Haeundae beach, South Korea. *Journal of Coastal Research*, (75): 1452–1456.

Mori, I.; de Silva, A.; Dusek, G.; Davis, J.; and Pang, A. 2022. Flow-based rip current detection and visualization. *IEEE Access*, 10: 6483–6495.

National Weather Service. 2025. Carolinas Rip Current Awareness. National Oceanic and Atmospheric Administration. Accessed: 2025-08-01.

Pitman, S.; Gallop, S. L.; Haigh, I. D.; Masselink, G.; and Ranasinghe, R. 2016. Wave breaking patterns control rip current flow regimes and surfzone retention. *Marine geology*, 382: 176–190.

Plyer, A.; Le Besnerais, G.; and Champagnat, F. 2016. Massively parallel Lucas Kanade optical flow for real-time video processing applications. *Journal of Real-Time Image Processing*, 11(4): 713–730.

Rampal, N.; Shand, T.; Wooler, A.; and Rautenbach, C. 2022. Interpretable deep learning applied to rip current detection and localization. *Remote Sensing*, 14(23): 6048.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6): 1137–1149.

Rombado, L.; Orescanin, M.; and Orescanin, M. 2024. Visualizing Bayesian Convolutional Neural Network Uncertainty In Coastal Images with Grad-Cam Ensembling. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, 1782–1785. IEEE.

Shen, S.; Kerofsky, L.; and Yogamani, S. 2023. Optical flow for autonomous driving: Applications, challenges and improvements. *arXiv preprint arXiv:2301.04422*.

Shepard, F.; Emery, K.; and La Fond, E. 1941. Rip currents: a process of geological importance. *The Journal of Geology*, 49(4): 337–369.

Shibata, M. 2023. Identifying risks for overseas-born beachgoers and suggesting future preventative strategies: a qualitative study based on interviews with 20 lifesavers from Australian tourist beaches. *Tourism in Marine Environments*, 18(1-2): 35–46.

Sohan, M.; Sai Ram, T.; and Rami Reddy, C. V. 2024. A review on yolov8 and its advancements. In *International Conference on Data Intelligence and Cognitive Informatics*, 529–545. Springer.

Tian, Y.; Ye, Q.; and Doermann, D. 2025. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*.

Wan, M.; Liu, T.; Bi, Y.; Wang, J.; Cui, H.; Cao, R.; Wang, J.; Shi, P.; Nie, N.; and Wang, Y. 2025. MCloudNet: An Ultra-Short-Term Photovoltaic Power Forecasting Framework With Multi-Layer Cloud Coverage. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 9908–9917.

Wang, H.-M.; Doong, D.-J.; and Lai, J.-W. 2025. Rip Current Identification in Optical Images Using Wavelet Transform. *Journal of Marine Science and Engineering*, 13(4): 707.

Wantzen, K. M.; Tharme, R.; and Pypaert, P. 2023. *River culture: life as a dance to the rhythm of the waters*. UNESCO.

Woodward, E.; Beaumont, E.; Russell, P.; and MacLeod, R. 2015. Public understanding and knowledge of rip currents and beach safety in the UK. *International Journal of Aquatic Research and Education*, 9(1): 49–69.

Yang, Q.; Wang, Y.; Liu, L.; and Zhang, X. 2024. Adaptive fractional-order multi-scale optimization TV-L1 optical flow algorithm. *Fractal and Fractional*, 8(4): 179.

Zach, C.; Pock, T.; and Bischof, H. 2007. A duality based approach for realtime tv-l 1 optical flow. In *Joint pattern recognition symposium*, 214–223. Springer.

Zheng, L.; Lu, W.; and Zhou, Q. 2023. Weather image-based short-term dense wind speed forecast with a ConvLSTM-LSTM deep learning model. *Building and Environment*, 239: 110446.