

Explainable Rip Current Detection and Visualization with XAI EigenCAM

Juno Choi¹, Muralidharan Rajendran², Yong Cheol Suh^{3,*}

¹ UR Interdisciplinary Program of Hydrography, Pukyong National University, Busan 48513, Korea.

² Research and Development Team, GreenBlue Inc., Busan 48934, Korea.

³ Department of Civil Engineering, Pukyong National University, Busan 48513, Korea.

jochoi@green-blue.co.kr, murali@green-blue.co.kr, suh@pknu.ac.kr

Abstract— Rip currents have long posed a serious threat to beachgoers and swimmers. Despite numerous preventive measures throughout the period, the fatality rate [1] underscores the need for a robust rip current detection system. Recently deep learning models have shown promising results in rip current detection, outperforming traditional methods. However, these models still exhibit some accuracy limitations due to insufficient data distribution. To address this challenge, we incorporate a novel largest dataset [2] comprising over 110,215 Korean coastline images. Through the comparative study of the state-of-the-art models, we aim to analyze the detection accuracy of each model and gain a deeper understanding of their intensity over rip current detection. In comparison to the other rip current datasets, the evaluation results on our proposed dataset demonstrate a remarkable elevation in accuracy, reaching 79.4 mAP. Further, we employ the EigenCAM (Eigen Class Activation Maps) to interpret the intense regions of the rip currents and to gain a deeper comprehension of rip current explainability. This comprehensive analysis marks a significant step toward improving rip current safety and understanding.

Keywords— Rip current detection, computer vision, class activation maps, largest rip current dataset, state-of-the-art models comparison.

I. INTRODUCTION

Rip currents are swift and narrow channels of water flowing from the shoreline out to the open sea. It presents a significant and potentially life-threatening hazard along coastlines worldwide. Rip currents have been responsible for numerous fatalities and near-drowning incidents. These currents exhibit variable strength, influenced by factors such as wave size, tide, and beach configuration. These factors render them unpredictable and challenging to identify through human visual observation. The water funneled through these channels can attain remarkable speed, reaching up to 2.43 ms^{-1} , which is faster than an average Olympic swimmer [3]. Predominant causes of rip currents are alterations in underwater topography such as sandbars and submerged depressions, and the configuration of breaking waves. Additionally, man-made structures such as piers and jetties can also contribute to the formation of rip currents. Traditional approaches to rip current detection primarily rely on human lifeguards for monitoring, yet accidents occur due to human errors and the unpredictability of rip currents. Consequently, an emergent need for automated

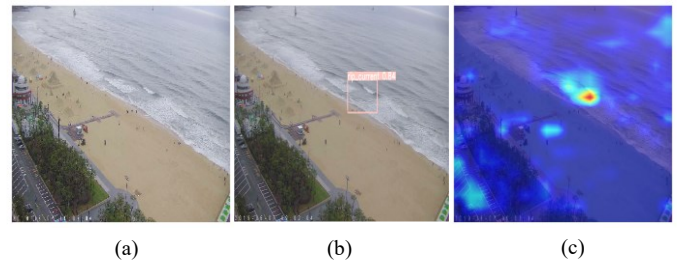


Figure 1. Illustration of predictions of visible rip currents on the coastline. Eigen-CAM method interprets the predicted rip currents on Haeundae Beach. (a) The original image, (b) Bounding box prediction of YOLOv8 model, (c) Visual explanations of Eigen-CAM over the same model and image.

rip current detection solutions has arisen. A subset of automated systems employs machine learning and numerical algorithms to scrutinize the complex data patterns associated with rip currents. These automated systems offer a notable advantage of continuous surveillance and expedited alerts, benefiting both beachgoers and lifeguards. Although these automated methods for rip current detection represent a marked improvement over traditional approaches, it is evident that further refinements and enhancements are necessary. In recent years, the coastal engineering community has leveraged deep neural networks to address numerous intricate challenges. Though deep learning models have proven successful in the realm of object detection techniques, there are several distinctive challenges exist in the case of rip current detection. Unlike traditional object detection scenarios like automobiles or buildings which have clearly defined object boundaries, rip currents are amorphous, lacking well-defined shapes or boundaries. These intrinsic complexities make rip currents detection exceptionally challenging.

One significant issue present in the literature of rip current detection is the absence of an appropriate dataset. This paper investigates the rip current detection on a specific coastline of the Republic of Korea and leverages an extensive dataset comprising over 110,215 images of Haeundae Beach, Busan. Every year a certain number of people get pulled away by these rip currents in Haeundae Beach. In 2017, a rip tide at this location carried over 70 persons 100 meters out to sea [4]. This distinctive dataset not only facilitates the training of detection models but also empowers the elucidation of rip current characteristics and explainability as the images are of a particular coastline. To make a suitable rip current preventive system, training the model and identifying the currents alone

are insufficient. It is crucial to not only build these models but also comprehend their predictions. However, with the rapid evolution of deep learning solutions, the models are becoming increasingly complex in terms of layers and parameters used. It is impossible to explicitly transform each computation into a logical explanation since the deep learning models are black-box solutions. Therefore, it is essential to utilize XAI tools to extract the information from different layers and transform it into comprehensible knowledge. In this paper, we employ the Eigen-CAM to explain the intensity of the rip current occurred regions. Furthermore, this study endeavors to conduct a comparative analysis of state-of-the-art object detection models concerning rip channels. This development of automated detection methodologies represents a significant stride in advancing beach safety measures. Additionally, it offers a more holistic and efficient approach to mitigating the persistent risks associated with rip currents. The major contributions of this paper can be summarized as:

- We leverage 110,215 region-specific coastline images both with and without rip current for the first time.
- We perform a comparative analysis with SoTA models of single-stage and two-stage detection models.
- We implement the XAI model explainability tool named Eigen-CAM to interpret the intensity of rip currents.
- We perform a careful data augmentation procedure to avoid over-fitting on specific region images.
- We demonstrate a high accuracy compared to other benchmark rip current detection datasets.

II. RELATED WORKS

Following the development of Machine Learning (ML) and Artificial Intelligence (AI), rip currents have become a well-investigated phenomenon. To meet the critical need for increased coastal safety, numerous studies have aided in building rip current detecting techniques. Traditional approaches to rip current identification have encompassed various methodologies such as marine radar systems for remote sensing observations [5], the deployment of dye tracking techniques [6] (as illustrated in Figure 2), and the integration of wave buoys [7]. Subsequently, researchers explored in-situ instrumental systems like Lifeguarding Operational Camera Kiosk System (LOCKS) [8], and current profiling meters [9]. These advancements have significantly enhanced precision and reliability. However, the prevailing challenge of these systems is the cost-effectiveness of implementing them on a widespread scale across various beaches and coastal regions.

The age of ML and AI began after the development of traditional numerical techniques. Maryan et al. [10] employed the Viola-Jones algorithm [11] and Convolution Neural Networks (CNN) for rip current detection using time-exposed images. Likewise, Rashid et al. [12] also used the same dataset as Maryan et al. but with a different rip current identification method. They proposed a Fully Convolutional Auto-Encoder architecture as RipNet. Later they proposed another network named RipDet [13] which is a more lightweight framework than RipNet as it utilizes the Tiny-YOLO architecture. Subsequently, the same authors introduced a new residual-

based architecture named RipDet+ [14] for enhanced feature detection on the same dataset as the prior version. However, the problem with these data and methodologies is that they are neither easily accessible nor easily understood by the average layman as a beachgoer. Hence, we use class activation maps to improve the explainability of the results of rip current detection.



Figure 2. Traditional dye tracking method for rip current identification [26].

de Silva et al. proposed a novel dataset [15], designed for detecting rip currents in individual frames. They opted for the Faster Region-based Convolutional Neural Networks (Faster RCNN) [16], which surpassed prior methods. Furthermore, they integrated a specialized temporal aggregation stage to enable detection from both still images and videos. Following this, de Silva introduced a novel feature detection model RipViz [17] aimed at extracting the essential features of rip currents from stationary videos. To gain insights into water flow behavior, they employed an LSTM autoencoder framework combined with pathline sequences. In the same way, Zhu et al. [18] significantly enhanced the performance by their YOLO-Rip model which is a lightweight network based on YOLOv5s [19] architecture. They incorporated the de Silva rip current dataset along with their collection of rip current images from the South China coastlines. However, some of the results of these studies exhibit false negative reports, due to the insufficient data distribution for training the model. We surpass this data distribution challenge by leveraging a novel rip current dataset with 110,215 coastline images. Further, we employ the Explainable AI tools to improve the explainability of the rip current detection. These combined enhancements significantly elevate the detection accuracy and ultimately safeguard beachgoers along coastal regions.

III. DATASET

The availability of meticulously annotated datasets remains a fundamental prerequisite for enhancing the efficacy of rip current detection. In the extant literature, several substantial datasets have been introduced for rip current detection. de Silva et al. [15] presented a novel dataset of 2440 images and 23 videos both with and without rip currents. A limitation of this dataset is that most of the images are of top-elevated aerial perspective, lacking diversity in terms of beach scene perspectives. In recognition of this limitation, Zhu et al. [18] supplemented an additional 1352 images to the de Silva dataset featuring varied perspectives. Similarly, Dumitriu et al. [20] leveraged the Zhu dataset and meticulously annotated 2466 rip current images for instance segmentation.

A. Proposed Dataset: Training Data

A substantial training dataset enables the extraction of intricate and nuanced patterns of an object in the image. The quantity of training images in a dataset directly correlates with the performance of the models. As the number of training images remains relatively modest in the existing literature, it potentially gives rise to suboptimal models with limited generalization and potential inaccuracies. To circumvent these problems and enhance the efficacy of rip current detection, we leverage a novel dataset [2] encompassing 110,215 coastline images of Haeundae Beach, Korea. It is 30x times larger than the Zhu et al. dataset which is the current largest rip current dataset with 3792 coastline images. It is a publicly available dataset that is collected by coastline CCTVs, marine observation buoys, and tide stations. It contains both rip current and non-rip current images and is meticulously refined through crowdsourced workforces. While there are other Korean coastlines rip current data also accessible on the same site, we specifically incorporate the Haeundae data as it has a larger number of images comparatively. The Haeundae rip current dataset features four distinct sets of data captured from varying camera perspectives and elevations, namely GLORY, PARA1, PARA2, and SEAC1. In total, our utilization comprises 89,181 coastline images depicting rip currents and 21,034 images devoid of rip currents. All these images are presented in High-Definition (HD) at dimensions of 1920×1080 pixels. The bounding box annotations are meticulously aligned with the x and y axes of an image, ensuring consistent reference points. The sample rip current images are shown in Figure 3.



Figure 3. Sample images of both with and without rip currents drawn from the proposed dataset. Images displayed in all four columns are of Haeundae Beach captured in different perspectives.

B. Test Data

A crucial phase in the evaluation of model effectiveness, accuracy, and generalization entails testing the data. Our testing data incorporates the rip current videos sourced from the test datasets of de Silva [15] and Dumitriu [20]. The de Silva dataset comprises 23 video clips, collectively comprising 18,042 frames, while the Dumitriu dataset features 17 videos comprising a total of 24,295 frames. Each testing video has a 20-second duration and a resolution of 1920×1080 pixels on average. The trained models underwent rigorous testing, encompassing a comprehensive assessment of their capabilities and performance.

IV. METHODOLOGY

Identifying rip currents in complex and turbulent ocean flow dynamics is undeniably challenging. While several alternative methods exist for rip current identification, deep learning approaches have emerged as the preeminent choice. In this section, we investigate several SoTA object detection models.

A. Faster RCNN

Faster R-CNN [16] is one among the family of region-based CNNs in the realm of computer vision. These object detection models typically comprise a shared feature extraction network alongside independent classification and localization networks. Faster R-CNN introduced the Region Proposal Network (RPN), a dedicated network that performs feature region extraction. RPN employed anchor boxes to extract region proposals more precisely. Furthermore, Faster R-CNN seamlessly integrated Fast R-CNN [21] as the final object detector in its architecture. This design enables the entire model process to execute seamlessly on the GPU, eliminating bottlenecks and enabling end-to-end network training.

B. YOLO v5

The You Only Look Once (YOLO) model series is a widely recognized family of object detection models, distinguished by its single-stage architecture. Among the series of models, the fifth version, the YOLOv5 model was developed by Jocher et al. from Ultralytics [19]. Notably, YOLOv5 exhibits faster convergence, requiring less time per epoch in comparison to its successors [22]. The architecture of YOLOv5 is delineated into three core components: the backbone, the neck, and the head. YOLOv5 employs a Cross Stage Partial Network (CSPNet) bottleneck [23] as the CNN backbone. The CSP models are based on DenseNet, a design aimed at reducing network parameters. These features are combined in the model neck and subsequently forwarded to the head. The model head interprets the combined features to predict the object classes in an image.

C. YOLO v8

A recent state-of-the-art version of the YOLO family is YOLOv8 [24] from the same creators of YOLOv5. YOLOv8 has emerged as a great choice for real-time object detection thanks to its updated architecture, refined convolutional layers in the backbone, and more sophisticated detection head and neck. The YOLOv8 model leverages the CSPDarknet-53 model as the backbone network, which significantly enhances both speed and precision compared to previous iterations. As an anchor-free model, it directly predicts the center of an object rather than its offset. This anchor-free paradigm effectively reduces the number of box predictions, consequently expediting the Non-Maximum Suppression (NMS) process. In the neck of the architecture, features are directly concatenated without necessitating uniform channel dimensions. This divergent architecture results in the reduction of parameter count and the overall size of the tensors. Similar to YOLOv5, YOLOv8 embraces online image augmentation during training. The mosaic data augmentation technique is consistently applied, affording the model exposure to objects in various positions.

V. EXPERIMENTS

This section encompasses the delineation of the experimental environment setup including hyper-parameter fine-tuning, computational resources utilized, the metrics employed for evaluation, and the utilization of the proposed dataset. In addition, we also investigate the methods of explainable AI with class activation maps to interpret the rip current detection results

A. Environmental Setup

The implementation of all baseline models outlined in the methodology section adheres to Python 3.10.9, PyTorch 2.0.1, and CUDA 11.8.0. The computational tasks, both training and inference, were executed on a dedicated GPU of NVIDIA GeForce RTX 3060 with 12GB of memory. The hyperparameters utilized in the foundational model were retained for training all the baseline models. Depending on the model size, training occurred with batches ranging from 16 to 64, spanning a duration of 100 epochs. Furthermore, we conducted experiments with varying threshold values to gain insights into their influence on the final detection performance.

B. Data Utilization

To facilitate the training process, the images are partitioned into a training set and a validation set. These images are accompanied by ground truth annotations and possess dimensions of 1920×1080 with three channels. For the sake of training efficiency, all images are uniformly resized to dimensions of 640×640 pixels while training the models. As all images of the dataset are of the same data distribution that is Haeundae Beach, we incorporate data augmentation techniques to prevent the model from overfitting to the original dataset and to enhance the diversity of the training dataset. The augmentation methods we utilized are as follows:

- **Rotation:** By applying random rotations to the images, models can learn about rip currents from different angles.
- **Flip:** Horizontal and vertical flipping help the model to recognize different orientations of rip currents.
- **Scaling:** Resizing images to scales such as 640×640 and 320×320 , allows the models to learn to identify rip currents of varying sizes.
- **Brightness and Contrast Adjustment:** This method aids the models in detecting rip currents under varying levels of illumination and lighting conditions.

C. Evaluation Metrics

To assess the performance and detection efficacy of the baseline models, we have employed four common evaluation metrics, namely Precision (P), Recall (R), mean Average Precision (mAP), and Frames Per Second (FPS). Precision and Recall are computed using the formulations presented in Equations (1) and (2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

In the context of our evaluation, TP (True Positive) represents the count of correctly detected rip currents, FP (False Positive) corresponds to the number of erroneously detected rip currents, and FN (False Negative) accounts for the missed rip currents. The primary evaluation metric for our model is mean Average Precision (mAP), with a focus on mAP50. This metric is based on the Intersection over Union (IoU), derived from the Jaccard index which signifies the similarity between two sets. Higher mAP values typically signify superior detection performance. The mean Average Precision is calculated as Equation (3), where k denotes the number of IoU thresholds and N denotes the total number of target categories.

$$\text{mAP} = \frac{\sum_{k=1}^{k=N} (AP_k)}{N} \quad (3)$$

Moreover, we assess the model's efficiency by considering training speed and inference time. In the evaluation of the test videos, we analyze the number of frames with precisely detected rip currents, employing the mAP50 metric to gauge the accuracy of each video while scrutinizing cases of failure. The detection speed is quantified in terms of Frames Per Second (FPS) which represents the number of images processed by the model in one second. Higher FPS values indicate faster image processing capability of the model.

D. Class Activation Mapping

The complexity of computer vision models can sometimes make it challenging to extract meaningful insights from their layers. Fortunately, there are XAI tools available to help make sense of these model layers. One such tool is Class Activation Maps (CAM), which combines activation maps from selected convolution layers to provide model explainability. These maps effectively visualize substantial features and are consistent with how humans perceive vision. CAM-based XAI methods

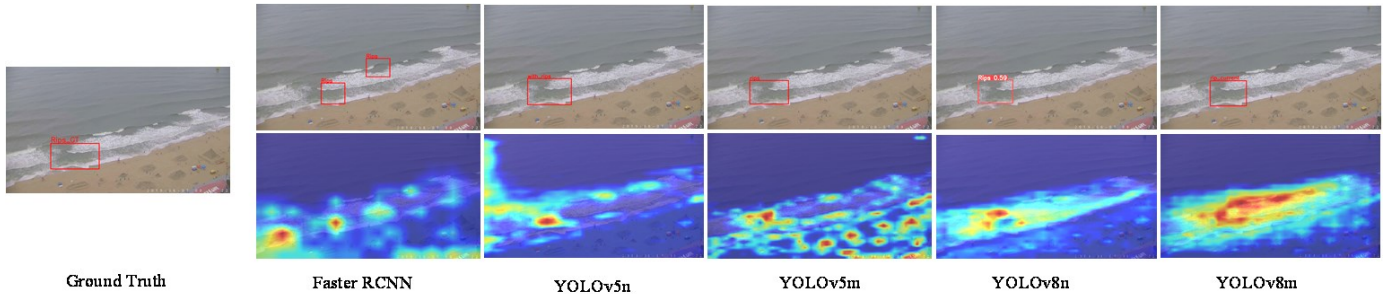


Figure 4. Visualization of obtained results of the state-of-the-art object detection models. For model comparison, same image has been used for inferencing. Each column represents different model outputs except Column 1 which shows the ground truth of the original image. While Row 1 displays the bounding box annotation predictions, Row 2 depicts the class activation map of the corresponding model.

provide insight into how the model is learning and whether the training setup needs any improvement. These methods use partial derivatives to weigh each activation map, making them faster than Region-based XAI methods. However, these saliency maps rely solely on feature maps and typically contain irrelevant features.

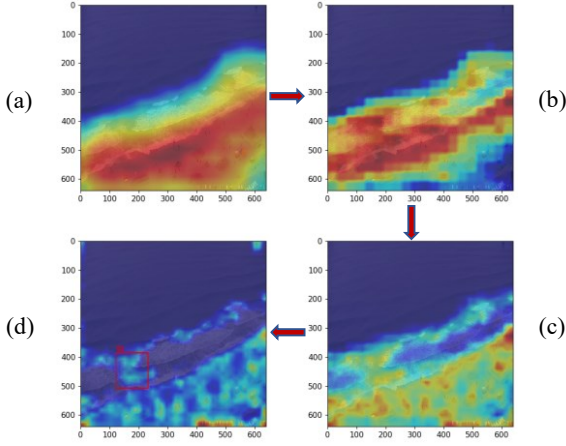


Figure 5. Different stages of activation maps from various layers of the model. Image (a) refers the activated map from bottom layers of the model where Image (d) refers the activated map from top layers of the model.

In this study, we incorporate the Eigen-CAM method, introduced by Muhammad et al. in [25]. Unlike other CAM approaches like Grad-CAM, Grad-CAM++, and XGrad-CAM, the Eigen-CAM stands out for its ease of implementation and seamless integration. It extends CAM by visualizing the principal components of learned features from various convolutional layers (as depicted in Figure 5). By overlaying this heatmap on the input image, the areas that generated the highest levels of activation from the convolutional layers can be emphasized (as shown in Figure 5(d)). It is important to note that as convolutional layers become more compact, they shift from acquiring more comprehensive aspects across the image to capturing tiny features. Consequently, the Eigen-CAM proposed using the Eigenvector and the primary weight matrix to create a heatmap. The n^{th} eigenvector next to the singular value decomposition generates various intriguing activation maps. As a result, this makes it possible to comprehend the amorphous boundaries of a detected rip current in the input image. The authors have expressed the Eigen-CAM as the equation below (Equation 4), where the class activation map, $L_{\text{Eigen-CAM}}$ is obtained by projecting $O_{L=k}$ onto the first eigenvector V_1 in the V matrix.

$$L_{\text{Eigen-CAM}} = O_{L=k} V_1 \quad (4)$$

VI. RESULTS

This section discusses the comprehensive analysis of the various object detection SoTA models and their associated performance on our dataset. We engage in a comparative analysis of single-stage object detection models from the YOLO series and a two-stage object detection model of Faster R-CNN. The performance of the models has been assessed both with and without the presence of rip currents.

Table 1. Evaluation loss comparison between the models.

Models	Box loss ↓		Object loss ↓	
	Train	Val	Train	Val
F-RCNN	1.5147	0.8806	1.3456	0.8880
YOLOv5n	0.0348	0.0305	0.0133	0.0146
YOLOv5m	0.9020	0.6651	0.4136	0.2808
YOLOv8n	0.0210	0.0072	0.0154	0.0060
YOLOv8m	0.8077	0.4044	0.2449	0.1218

As Table 1 provides insights into various losses over training and validation datasets, the YOLOv8 nano model has lesser loss values compared to other models. The rip currents detected by the trained models on our dataset are visually depicted in Figure 4. It's imperative that the models should accurately detect rip currents from the given data. However, as the rip currents lack clearly defined boundaries, it causes a significant amount of uncertainty in the scale of actual ground truth bounding boxes [15]. Accordingly, to classify a detection as accurate, the detected bounding box need not have a striking resemblance to the labeled bounding box. Given these characteristics, we adopt a lower IoU threshold for the detection. The Faster RCNN model has demonstrated expected results in detecting rip currents as shown in Figure 4. However, as the RCNN models operate as a two-stage detection algorithm, they encounter notable challenges as the dataset size increases. While the size of the dataset grew, the detection time increased substantially. Timely detection of rip currents holds paramount importance as it directly impacts the safety of swimmers and beachgoers. In response to these challenges, we have explored the capabilities of the state-of-the-art YOLO (You Only Look Once) detection frameworks. YOLO offers distinct advantages compared to the Faster RCNN network, which aligns with the real-time demands of rip current detection, further enhancing beach safety. The advantages of YOLOv5 are more pronounced as illustrated in Tables 1 and 2, where it surpasses Faster RCNN. Notably, the YOLOv5 nano-size model performs better than the YOLOv5 medium-size model.

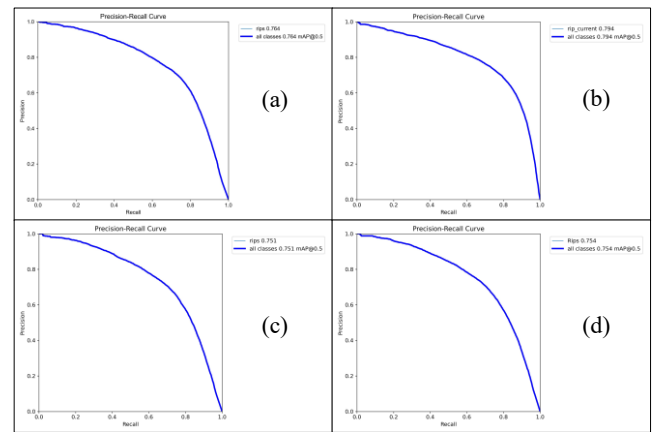


Figure 6. Plots showing the mean Average Precision values of different models with different sizes. In this PR curve plot, while the x axis denotes recall, y axis refers precision. Plot (a) denotes the value of YOLOv5n, Plot (b) denotes the value of YOLOv8n, Plot (c) denotes the value of YOLOv5m, and Plot (d) denotes the value of YOLOv8m.

Table 2. Experimental results of the object detection SoTA models.

Models	Precision	Recall	mAP50	mAP50:95
F-RCNN	0.6434	0.6457	0.6645	0.3482
YOLOv5n	0.7047	0.7271	0.7637	0.4147
YOLOv5m	0.7035	0.7041	0.7512	0.4127
YOLOv8n	0.7139	0.7683	0.7940	0.4259
YOLOv8m	0.7074	0.7004	0.7536	0.4159

The latest model YOLOv8 demonstrates exceptional capabilities in rip current detection, even when multiple rip currents are present. In our comprehensive analysis, the YOLOv8 outperforms the Faster RCNN and YOLOv5 models. As shown in Table 2, the YOLOv8n model attains an overall mAP score of 0.794. Furthermore, the YOLOv8 model size is 14.6 MB, which underscores the compact model size in comparison to other models. This feature is crucial when considering using the model in a coastal early warning system, as a smaller model size has advantages.

VII. CONCLUSION

We introduced a novel dataset for rip current detection, setting a new benchmark in terms of dataset scale. Employing state-of-the-art models including Faster-RCNN, YOLOv5, and YOLOv8, we conducted a comprehensive comparative analysis encompassing precision, recall, mAP50, and mAP50:95 performance metrics. Among other models, YOLOv8 emerged as the most effective model and demonstrated its suitability for detecting the amorphous nature of rip currents. Additionally, we explored the model explainability tools using the Explainable AI method namely Eigen-CAM. The activation maps interpret the black box model computations at various stages and convert them into comprehensible knowledge. Our future research endeavors are focused on expanding the dataset to encompass a wider array of rip current scenarios, exploring advanced model frameworks, and developing real-time rip current detection systems designed for coastal CCTV cameras. Furthermore, adding additional information such as wave patterns, lighting conditions, and bathymetrical details might also be beneficial in lowering false-negative detections.

REFERENCES

- [1] N. N. W. S. US Department of Commerce, "Surf Zone Fatalities in the United States in 2023: 87".
- [2] "AI-Hub." Accessed: Sep. 11, 2023. [Online]. Available: <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realml&dataSetSn=71297>
- [3] N. O. and A. A. US Department of Commerce, "Rip Currents - Currents: NOAA's National Ocean Service Education".
- [4] "Haeundae riptide sweeps more than 70 people out to sea - The Korea Times." Accessed: Oct. 06, 2023. [Online]. Available: https://www.koreatimes.co.kr/www/nation/2023/10/113_233993.html
- [5] M. C. Haller, D. Honegger, and P. A. Catalan, "Rip Current Observations via Marine Radar," *J Waterw Port Coast Ocean Eng*, vol. 140, no. 2, pp. 115–124, Mar. 2014, doi: 10.1061/(ASCE)WW.1943-5460.0000229.
- [6] H. D. Kim, K.-H. Kim, H. D.; Kim, K.-H. Kim, H. Kowalewska-Kalkowska, and A. Cedro, "Analysis of Rip Current Characteristics Using Dye Tracking Method," *Atmosphere* 2021, Vol. 12, Page 719, vol. 12, no. 6, p. 719, Jun. 2021, doi: 10.3390/ATMOS12060719.
- [7] B. K. Haus, "Remote sensing applied to rip current forecasts and identification," *Chapter*, vol. 8, pp. 133–145, 2011.
- [8] Y. Liu and C. H. Wu, "Lifeguarding Operational Camera Kiosk System (LOCKS) for flash rip warning: Development and application," *Coastal Engineering*, vol. 152, p. 103537, Oct. 2019, doi: 10.1016/J.COASTALENG.2019.103537.
- [9] J. MacMahan *et al.*, "An Introduction to Rip Currents Based on Field Observations," <https://doi.org/10.2112/JCOASTRES-D-11-00024.1>, vol. 27, no. 4, Jul. 2011, doi: 10.2112/JCOASTRES-D-11-00024.1.
- [10] C. Maryan, M. T. Hoque, C. Michael, E. Ioup, and M. Abdelguerfi, "Machine learning applications in detecting rip channels from images," *Appl Soft Comput*, vol. 78, pp. 84–93, May 2019, doi: 10.1016/J.ASOC.2019.02.017.
- [11] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int J Comput Vis*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISL.0000013087.49260.FB/METRIS.
- [12] A. H. Rashid, I. Razzak, M. Tanveer, and A. Robles-Kelly, "RipNet: A Lightweight One-Class Deep Neural Network for the Identification of RIP Currents," *Communications in Computer and Information Science*, vol. 1333, pp. 172–179, 2020, doi: 10.1007/978-3-030-63823-8_21/COVER.
- [13] A. H. Rashid, I. Razzak, M. Tanveer, and A. Robles-Kelly, "RipDet: A Fast and Lightweight Deep Neural Network for Rip Currents Detection," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2021-July, Jul. 2021, doi: 10.1109/IJCNN52387.2021.9533849.
- [14] A. H. Rashid, I. Razzak, M. Tanveer, and M. Hobbs, "Reducing rip current drowning: An improved residual based lightweight deep architecture for rip detection," *ISA Trans*, vol. 132, pp. 199–207, Jan. 2023, doi: 10.1016/J.ISATRA.2022.05.015.
- [15] A. de Silva, I. Mori, G. Dusek, J. Davis, and A. Pang, "Automated rip current detection with region based convolutional neural networks," *Coastal Engineering*, vol. 166, p. 103859, Jun. 2021, doi: 10.1016/J.COASTALENG.2021.103859.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Adv Neural Inf Process Syst*, vol. 28, 2015, Accessed: Sep. 08, 2023. [Online]. Available: <https://github.com/>
- [17] A. de Silva *et al.*, "RipViz: Finding Rip Currents by Learning Pathline Behavior," *IEEE Trans Vis Comput Graph*, 2023, doi: 10.1109/TVCG.2023.3243834.
- [18] D. Zhu, R. Qi, P. Hu, Q. Su, X. Qin, and Z. Li, "YOLO-Rip: A modified lightweight network for Rip currents detection," *Front Mar Sci*, vol. 9, p. 930478, Aug. 2022, doi: 10.3389/FMARS.2022.930478/BIBTEX.
- [19] G. Jocher *et al.*, "ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation," *Zenodo*, 2022.
- [20] A. Dumitriu, F. Tatui, F. Miron, R. T. Ionescu, and R. Timofte, "Rip Current Segmentation: A Novel Benchmark and YOLOv8 Baseline Results," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1261–1271, Jun. 2023, doi: 10.1109/CVPRW59228.2023.00133.
- [21] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, 2015.
- [22] M. Durve *et al.*, "Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications," *The European Physical Journal E* 2023 46:5, vol. 46, no. 5, pp. 1–7, May 2023, doi: 10.1140/EPJE/S10189-023-00290-X.
- [23] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," Nov. 2019.
- [24] "YOLOv8 - Ultralytics | Revolutionizing the World of Vision AI." Accessed: Sep. 20, 2023. [Online]. Available: <https://ultralytics.com/yolov8>
- [25] M. B. Muhammad and M. Yeasin, "Eigen-CAM: Class Activation Map using Principal Components," *Proceedings of the International Joint Conference on Neural Networks*, Jul. 2020, doi: 10.1109/IJCNN48605.2020.9206626.
- [26] N. N. W. S. US Department of Commerce, "Rip Current Photos".