Automated Rip Current Detection with Region based Convolutional Neural Networks

Akila de Silva[a], Issei Mori[a], Gregory Dusek[b], James Davis[a] and Alex Pang[a]

[a]*University of California, Santa Cruz, CA, United States*
[b]*NOAA National Ocean Service, Silver Spring, MD, United States*

ARTICLE INFO

ABSTRACT

This paper presents a machine learning approach for the automatic identification of rip currents with breaking waves. Rip currents are dangerous fast moving currents of water that result in many deaths by sweeping people out to sea. Most people do not know how to recognize rip currents in order to avoid them. Furthermore, efforts to forecast rip currents are hindered by lack of observations to help train and validate hazard models. The presence of web cams and smart phones have made video and still imagery of the coast ubiquitous and provide a potential source of rip current observations. These same devices could aid public awareness of the presence of rip currents. What is lacking is a method to detect the presence or absence of rip currents from coastal imagery. This paper provides expert labeled training and test data sets for rip currents. We use Faster-RCNN and a custom temporal aggregation stage to make detections from still images or videos with higher measured accuracy than both humans and other methods of rip current detection previously reported in the literature.

1. Introduction

Rip currents are the most significant safety risk to swimmers along the coastlines of oceans, seas, and large lakes. [8, 7, 13]. The majority of beach goers do not know how to identify rip currents, and there is no robust and reliable location-independent method to identify them. Globally there are thousands of drownings each year due to rip currents [23, 42]. A 20 year study by the US Lifesaving Association reports that 81.9% of the 37,000 beach rescues each year are due to rip currents [7]. There has been no decline in the number of associated drowning fatalities, despite warning signs and educational material.

Rip currents are a well-studied ocean phenomenon [3, 29, 31]. They are defined as strong and narrow channels of fast-moving water that flow towards the sea from beaches. When waves break, they form a "setup" or an increase in mean water level. This setup can vary along a shoreline depending on the amount or height of breaking waves. Rip currents form as water tends to flow alongshore from regions of high setup (larger waves) to regions of lower setup (smaller waves) where currents converge to form a seaward flowing rip. The speed of seaward rips can be quite strong reaching 2 m/s, faster than an Olympic swimmer. There are multiple factors that determine the location and strength of rips, such as bathymetry, wave height and direction, tide, and beach shape. Rip currents may either be transient or persistent in space and time. Rips that are frequently found at the same location are usually indicative of a fairly stable bathymetric feature such as a sand bar or reef, or a hard structure such as rocky outcrop, jetty or pier. These bathymetric features results in variations in wave breaking and setup leading to channelized rip current flow. Transient or flash rips are independent of bathymetry and may move up or down the beach, and may appear or disappear.

ORCID(s):

Lifeguards are often trained to identify rip currents. However the majority of drownings occur on beaches without trained personnel [1, 4]. Posted signs can provide a warning, but there is evidence that most people do not find existing signs helpful in actually identifying rip currents [6].

Experts at the National Oceanic and Atmospheric Administration (NOAA) use images and video to gather statistics about rip currents [19]. These data are supporting the validation of a rip current forecast model to alert people to potential hazards [20]. The most commonly used method to visualize rip currents from video is time averaging, summarizing a video as a single image [28]. In [43] boosted cascade of simple Haar like features, a machine learning technique, was used to detect rip currents in time averaged images. However these time averages when manually assessed can be misinterpreted. Furthermore, they are not readily available nor interpretable by the average beachgoers, and the process of averaging removes available information.

In recent years the coastal engineering community has successfully used deep neural networks to solve many problems. Classification problems such as classifying wave breaking in infrared imagery [9], beach scene and other landscape classification [11], automated plankton image classification [41] *and* ocean front recognition [36] were formulated as deep learning problems using convolutional neural networks. Furthermore, some regression problems such as optical wave gauging [10], tracking remotely sensed waves [52], typhoon forecasting [33] were also solved using deep neural networks. In addition, generative adversarial networks, a type of deep neural networks, were used to improve the quality of downscaling of ocean remote sensing data [18].

Object detection with deep neural networks is well studied in the computer vision community. However most benchmarks and research focus on detecting physical objects with boundaries between what is and is not an object [16, 22, 37]. Rip currents are ephemeral "objects"

Figure 1: A collection of beach scenes, some of which contain dangerous rip currents. Unfortunately these "objects" do not have clear shape, and most people find them hard to identify. This paper describes a computer vision system with detection accuracy higher than both existing published methods and human observers.

which are not observable in every frame, and amorphous without clearly defined boundaries even when observable. It is not obvious whether existing methods are applicable. Figure 1 provides a set of examples, illustrating the difficulty of the problem.

Our work is aimed at introducing this problem to the coastal engineering community, and showing that object detection methods *are* applicable. We gathered training data sets of rip currents and worked with experts at NOAA to ensure that test data were labeled correctly. We use Faster-RCNN [50] with a custom temporal aggregation stage that allowed us to achieve detection accuracy that is higher than both humans and other methods of rip current detection previously reported in the literature.

The contributions of this paper are:

- Evidence that region based convolutional neural networks (CNN) approach for object detection is applicable to amorphous and ephemeral objects such as rip currents

- Analysis showing rip current detection accuracy above existing
  published methods

- Data sets of rip current images and video for training and testing

The remainder of the paper is organized as follows. All the related work is summarized in Section 2. We discuss how the data was collected in Section 3. Our method is discussed in 4. Results are discussed in 5. Limitations and discussion are in Section 6. In Section 7 we conclude our paper. And in Appendix A we provide the link to the supplementary materials.

2. Related Work

Rip currents are most commonly studied using in situ techniques or instrumentation. Fluorescein dye is commonly released into the ocean and the dispersion observed [5, 14, 15, 48]. Wave sensors, acoustic velocimeters, or current profilers can be deployed at specific locations [21, 32, 34, 40]. Floating drifters with embedded GPS units have also been used to measure currents [12, 13, 51]. These methods are costly, time consuming, require technical expertise and are generally only applicable to highly localized instances in time and space. These limitations severely hinder the applicability of such approaches to both public warnings and model validation.

Time averaged images are a routine method for analyzing video in oceanic research, with 10 minutes being a common integration period [27, 28, 38, 47]. This method is popular because averages often make identification of rip channels easier for the human eye. While these images are usually intended for human interpretation, Maryan et. al. apply machine learning to recognize rip channels in time averaged images [43]. Nelko also used time averaged images and noted that prediction schemes developed at one beach location may not be directly applicable to another without some modifications [44]. In contrast to these works, the analysis in this paper suggests that object detection on individual frames outperforms time averaged images.

Dense optical flow [2, 30] has been used to detect rip currents in video [46]. This method is attractive since optical flow fields can be directly compared against ground truth flow fields obtained from in situ measurements [17]. Unfortunately these methods are sensitive to camera perturbation, and have difficulty in areas lacking textural information. The results in this paper suggests that object detection on individual frames outperforms previous optical flow based methods.

Certain kinds of rip currents are characterized by visible sediment plumes. These can be segmented based on changes in coloration. For example, Liu et al, use thresholding in HSV color space to detect rip currents [40]. Unfortunately, not all rip currents contain sediment plumes, and thus this method is not applicable to our data sets.
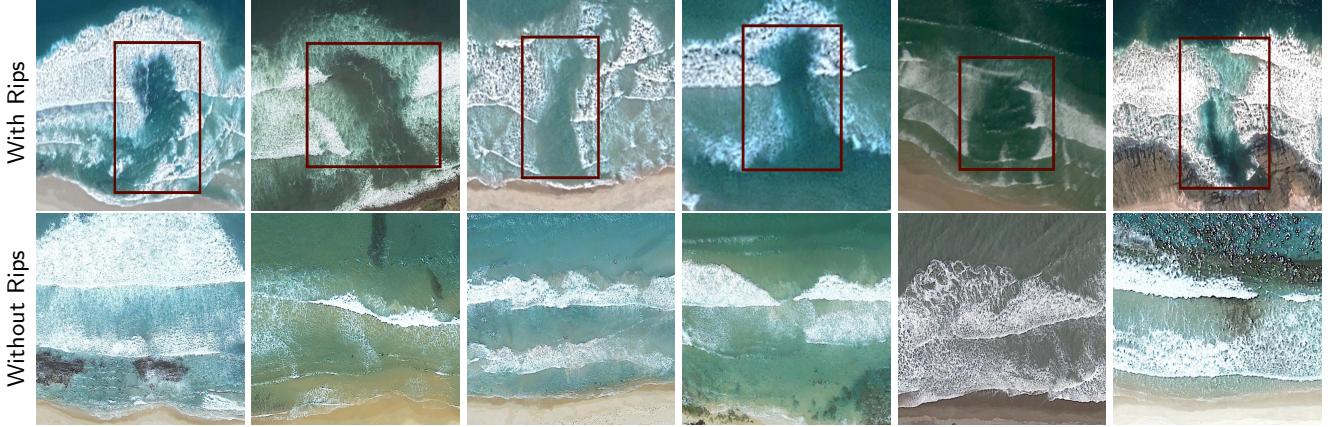
Object detection in images is well studied in the com-

Figure 2: Examples drawn from the 2440 images we collected and labeled to build a training data set. Ground truth bounding boxes are shown in red.

puter vision literature [25, 39, 45, 49, 53]. These methods have been extended to detect objects in videos [26, 35, 54]. This work has not previously been applied to rip currents both because it is not clear there is an "object" to detect, and because of a lack of publicly available data sets for training and testing.

This paper contributes labeled data sets for rip current detection, and evidence that object detection outperforms existing published methods on this application.

3. Data sets

3.1. Training Data Images

Since rip currents are a new problem domain for computer vision, we did not find any existing public databases of rip current images. Therefore, we assembled a training data set of rip current images and non-rip current images from scratch. Our primary source for the database was Google Earth^{TM}, which allowed us to extract high-resolution aerial images of rip currents and non-rip currents. In total, the database contains 1740 images of rip currents and 700 images of similar beach scenes without rip currents. The images range in size from 1086×916 to 234×234 . We annotated ground truth in the rip current images with axis-aligned bounding boxes. Some examples of the training data set are shown in Figure 2. Note that this data set contains unambiguous easy examples. We used this static image data set for training models described in Section 4.

3.2. Test Videos

We also collected a data set of 23 video clips consisting of 18042 frames. There are a total of 9053 frames with and 8989 frames without rip currents. Image size varies from 1280×720 to 1080×920 . Ground truth annotations were verified by a co-author who is also a rip current expert at NOAA. Figure 1 contains both positive and negative example frames, as well as a few frames from the training set that might be mistaken as containing a rip current. Note that these sequences contain more difficult cases. The rip currents are not visible in every frame, but a positive bounding

box is applied whenever viewing the entire video segment indicates that a rip current is present.

The frames of this video data set were used for testing. Note that the static images in the training set were taken from high elevation while the videos used to test the model were taken from a lower perspective. Even so, the trained model performed well on the test frames from the video collection.

In order to encourage progress in this domain, both data sets will be made available to the public. We are including a link to the data set in the supplementary materials in appendix A.

4. Method

4.1. Static Image Detector: Faster RCNN

Region-based convolutional neural networks have achieved great success in object detection problems. These object detection models usually consist of separate classification and localization networks with a shared feature extraction network. We use Faster RCNN which is also composed of two components. The first is the deep convolutional neural network that proposes regions. The second is the Fast RCNN detector [24].

Faster RCNN follows the traditional object detection pipeline. It first generates region proposals, and then categorizes each proposal as either rip current or background. Secondly, the classified bounding boxes are further refined. Essentially, the model learns a mapping from the generated regions to the actual ground truth with a regression network. The model then uses this mapping "function" during testing to refine the generated region. These refined bounding boxes can be anywhere in a frame as features are translation invariant [50]. If there is more than one bounding box detected in a frame, we only keep track of the largest one and ignore any additional boxes.

We trained the Faster RCNN model with the static image data set. Before training, each image was augmented by rotating 90° degrees clockwise and counter clockwise, producing a training data set three times the size of the original training data set. All the training data was resized to
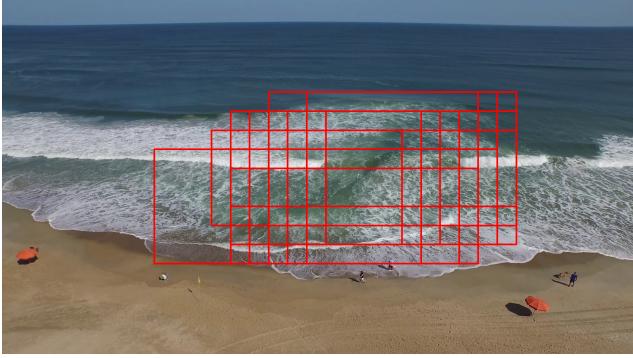
Figure 3: Amorphous phenomena without clear boundaries, like rip currents, result in wide variation in detected bounding boxes. All detected bounding boxes from the frames of one video are superimposed onto a single image. Bounding boxes may differ significantly even in consecutive frames.

Figure 4: Frame aggregation for a window of length N. First row shows the input frame sequence. Second row shows the detections from Faster RCNN. Third row shows the accumulation buffer.

300×300 before training to save computation time.

4.2. Frame Aggregation

Static object detection models only consider the information in the frame currently being processed. However, rip currents are natural ocean phenomena, with shape and texture change depending on many external factors such as weather, wind speed, wave field characteristics, water flow speed, floating debris, and dirt sediments. The exact boundaries of a rip current are not well defined. This is different than objects with well-defined edges such as pedestrians or vehicles. Applying detection algorithms to objects with amorphous boundaries such as rip currents produces bounding boxes with variable sizes and locations in adjacent video frames. In Figure 3 we illustrate this variability by drawing all correctly detected bounding boxes from one video sequence onto a single frame.

This variability affects overall accuracy, and would not instill confidence in the results if these bounding boxes were presented to a user as a video overlay. Thus we investigate temporal smoothing and aggregation to improve the results.

We find the overlapping regions of the detected bounding boxes by using an accumulation buffer with the same size as the input frame and initialized as a zero matrix. We consider a temporal window of N frames to build the accumulation buffer. In the first $N - 1$ frames, the accumulation buffer is incremented by 1 for each region within a detection bounding box. Starting with frame N, the area covered by a detection bounding box is incremented by 1 and capped at a maximum of N. Regions not covered by a detection bounding box are decremented by 1, but retains a minimum of 0 in the accumulation buffer. In effect, the accumulation buffer keeps track of the bounding boxes using a sliding window of N frames. The process of building the accumulation buffer is illustrated in Figure 4, where areas with higher values are displayed as darker regions in the accumulation buffer. For purposes of identifying a single bounding box over the collection of bounding boxes across N frames, we
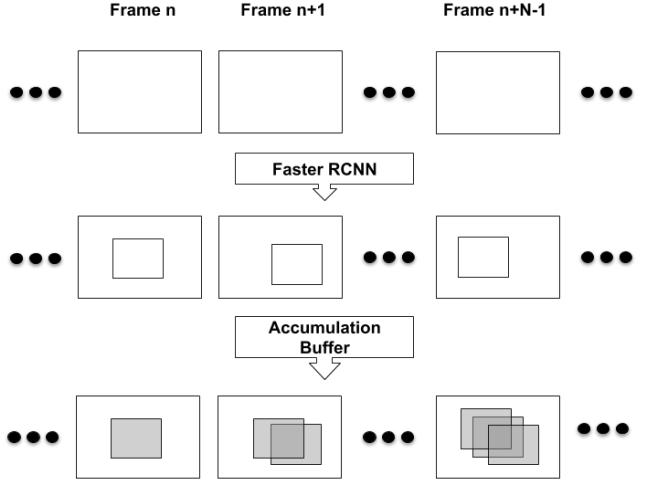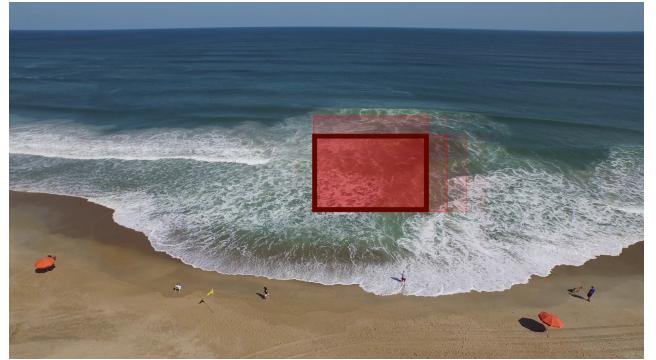

Figure 5: A visualization of the resulting accumulation buffer values. Regions with a higher value are shown in more opaque red. The resulting bounding box around thresholded values is shown in solid dark red. The full video can be seen in the supplementary material in appendix A

consider only the parts of the accumulation buffer where the value is at least T, and draw the tightest possible axis-aligned bounding box around this region (see Figure 5). This is the aggregated detection. In our implementation we use N=60 and T=50.

Before frame aggregation, large variations in bounding box size occur in almost all consecutive frames, shown in Figure 6 top. After frame aggregation, the average size change is much smaller, with most frames having zero change in size from the prior frame. The variation in position of the bounding boxes is similarly reduced by frame aggregation, as seen in Figure 6 bottom. This improved temporal coherence provides a smoother and more consistent portrayal of the rip current extent when shown as an overlay on the video.

.	Human	Philip [46]	Maryan [43]	Maryan [modified]	F-RCNN [ours]	F-RCNN+FA [ours]
rip_01.mp4	0.976	0.895	0.358	0.460	0.966	1.000
rip_02.mp4	0.700	0.098	0.698	0.135	0.776	0.860
rip_03.mp4	0.231	0.347	0.194	0.540	0.831	0.950
rip_04.mp4	0.757	0.800	0.487	0.780	0.939	0.970
rip_05.mp4	0.883	0.280	0.736	0.450	0.834	0.957
rip_06.mp4	0.881	0.000	0.167	0.470	0.753	0.890
rip_08.mp4	0.492	0.063	0.328	0.730	0.860	0.850
rip_11.mp4	0.824	0.000	0.563	0.940	0.930	0.951
rip_12.mp4	1.000	0.000	0.734	1.000	1.000	1.000
rip_15.mp4	0.967	0.137	0.315	0.390	0.760	0.870
rip_16.mp4	0.614	0.073	0.468	0.640	0.820	0.920
rip_17.mp4	1.000	0.321	0.064	0.750	0.980	1.000
rip_18.mp4	0.563	0.218	0.250	0.240	0.790	0.890
rip_21.mp4	0.901	0.543	0.486	0.180	0.940	1.000
rip_22.mp4	0.583	0.000	0.479	0.395	0.880	0.974
no_rip_01.mp4	0.986	0.169	0.000	0.972	0.813	1.000
no_rip_02.mp4	1.000	0.789	0.000	0.985	0.807	1.000
no_rip_03.mp4	0.919	0.000	0.000	0.981	0.984	1.000
no_rip_04.mp4	0.952	0.000	0.000	0.974	0.835	1.000
no_rip_05.mp4	0.903	0.000	0.000	0.986	0.833	1.000
no_rip_06.mp4	1.000	0.246	0.000	0.986	0.875	1.000
no_rip_07.mp4	0.983	0.525	0.000	0.982	0.875	1.000
no_rip_11.mp4	0.988	0.198	0.000	0.964	0.924	1.000
average accuracy	0.760	0.307	0.210	0.729	0.884	**0.984**

Table 1
Accuracy for each video in the test set. Column 3: Maryan[43] is trained on their training data. Column 4: Maryan [modified] is trained on our training data. Our method has higher overall accuracy than humans or any of the prior methods tested. Frame aggregation does contribute to improvement in accuracy.
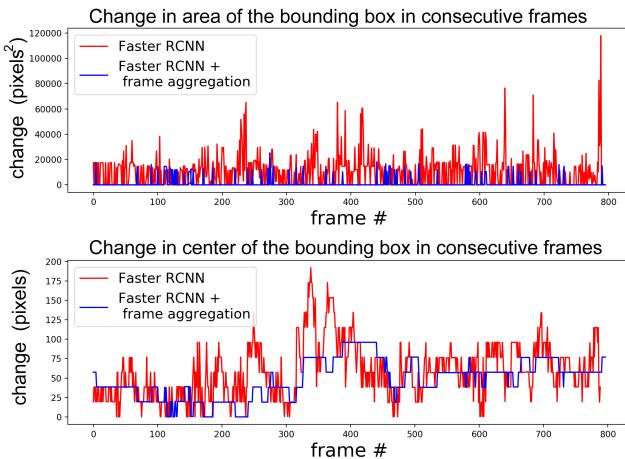

Figure 6: Plots showing the differences in area and center of bounding boxes in consecutive frames. Without frame aggregation (in red) there are much higher differences in bounding box sizes and positions than there is after frame aggregation (in blue).

5. Results

We compared the accuracy of our method with human observers as well as two prior methods. We also compared our own method with and without temporal aggregation.

Comparison metric. All methods were tested using the video data set. Frames were labeled as correctly classified if the detected bounding boxes have an Intersection over Union (IoU) score versus ground truth above 0.3. IoU is calculated as *area_of_intersection/area_of_union* of the ground truth and the detected bounding boxes. Accuracy of the video was computed as *correct_labels/total_frames*, and Table 1 provides the results for all methods.

Humans. The primary reason for an automated method of rip current detection is that most people are not good at identifying rip currents [6]. Human annotators were asked to draw bounding boxes around places they believe to have rip currents. We sampled approximately every tenth frame from our video test set and randomized presentation order across all positive and negative examples. Human annotators were not carefully trained, instead they were provided three positive and three negative examples, roughly the amount of information which might fit on a sign at the beach. Annotators were acquired using Mechanical Turk, with basic screening for reliable workers, and paid $0.10 per image.

Although human performance was relatively poor with only 76% of frames labeled correctly, it was higher than we expected based on past studies [6]. We hypothesize that the clearly labeled sample images on a web page directly adja-

Figure 7: Rip current detections on some of the frames from the test data set. Bounding boxes show the correctly detected rip currents. Frames without bounding boxes do not contain any rip currents.

cent to the image they were asked to label was a positive contributing factor.

Time averaged images. Maryan et al. [43] perform detection on time averaged images using a boosted cascade of simple Haar like features [53]. We used Maryan's time averaged data for training since our training data set consists of only static frames. Testing was performed by first computing time averaged images on each video in our test data set. This method did not perform well. In order to determine if the cause was the images available for training or the model itself, we repeated the experiment with new data. We replaced the relatively small number of low resolution time-averaged images from [43] with the static images from our training data set. Testing this time was against single frames in our test videos. This modification is called Maryan[modified] in the results table. When using our training data the model accuracy improved considerably, leading us to conclude that appropriate training data is critical to good results. Furthermore, it suggests further investigation on whether using crisp images, rather than time averaged images, to train a model might produce more accurate results.

In our test images/videos the beach is always located at the bottom half of the image/video. However, the training data used by [43] was cropped from images where the beach is located in the top half of the image. Therefore, we were

concerned that maybe the difference of orientation between the training data and the test data contributed to the low accuracy of the model. However, when we retrained the model with vertically flipped training data we did not see any significant difference in accuracy on our test images/videos. We hypothesize that the reason for this is that the cropped training images contain insignificant amount of beach pixels.

Optical flow. Philip et al. [46] compute optical flow on video sequences and make the simplifying assumption that rip currents can be identified by regions with the second most predominant flow direction, after that of the primary incoming wave direction, and that they flow in a single seaward direction. This results in regions of actual rip currents, but also picks up swash regions where water is washed up the beach and back out to sea with the passing of each wave. This method was introduced with the primary intention of providing visualizations to users, rather than automated detection. To allow comparison, we modify the method to return a bounding box around the largest detected region, ignoring smaller regions which are less likely to be correct. This method performed poorly on our test videos. We noticed that in videos where there is not enough textural information on the rip current, the optical flow field generated was weak, leading to either missed detections or detection in other regions of the video with stronger texture.

Frame aggregation. We implemented frame aggregation as a post process to Faster RCNN initially to temporally stabilize detections, driven by a need for user interpretable visualization of rip current location.

In order to understand whether temporal smoothing also increased accuracy we analyzed our implementation both with and without frame aggregation. We found that temporal aggregation leads to higher accuracy than using Faster RCNN alone. Example detection results are shown in Figure 7. Numerical comparison of humans, prior methods, and our model are provided in Table 1. Faster RCNN with frame aggregation had the highest accuracy in nearly all cases, and the highest overall (last column of Table 1). For visual comparisons we have added all the results in the supplementary materials at appendix A.

6. Discussion and Future Work

As with all machine learning models, our implementation can fail when used with images that do not resemble the training data set. Our data sets included primarily rip currents characterized by a gap in breaking waves, the most common visual indicator for bathymetry controlled rip currents. Thus we would expect to miss rip currents with other visual indicators like sediment plumes. We also expect to fail when presented with new imagery, and occasionally for no apparent reason at all, as seen in Figure 8.

We found it difficult to compare our method with prior work and verify that our model performs well in all conditions previously researched, due to a lack of public data sets on which to verify our results. In order to ensure that future work has a baseline from which to compare, our data sets with thousands of labeled frames will be made public. Nevertheless our data sets are still limited. The accuracy numbers presented in this paper are correct on this limited data, but almost certainly overstate probable outcomes in real world deployment. We expect that future work will need to collect more examples including less common rip current visual presentation, a greater variety of scales, and a wider array of beach distractors.

Lastly, our work lacks a success metric which is meaningful to real users. Certainly IoU, true positive rate, mAP, and the like are common in computer vision research, but is it appropriate to measure accuracy on single frames? Most conceivable deployed systems, such as pointing a mobile phone at the beach, have access to video. If measuring accuracy on video, should we measure accuracy aggregated over 1 second or 1 minute? Is a tight bounding box measured by IoU needed, or just a general region? A metric useful to researchers and simultaneously meaningful to users is needed.

7. Conclusion

We present a machine learning approach for identifying rip currents automatically. We use Faster RCNN and a custom temporal aggregation stage to make detections from still images or videos with higher measured accuracy than both

Figure 8: Example failure cases. The false positives on the beach scene (right) are not easily explainable. The false positive on the left scene happens only on spurious frames, which is then corrected by frame aggregation.

humans and other methods of rip current detection previously reported in the literature. Training data set and suite of test videos are made available for other researchers.

A. Appendix

Supplementary material to this article can be found online at https://sites.google.com/view/ripcurrentdetection/home

References

[1] Australia, S.L.S., 2019. National coastal safety report. https://issuu.com/surflifesavingaustralia/docs/ncsr2019 .

[2] Barron, J.L., Fleet, D.J., Beauchemin, S.S., 1994. Performance of optical flow techniques. International Journal of Computer Vision 12, 43–77. URL: https://doi.org/10.1007/BF01420984, doi:10.1007/BF01420984.

[3] Bowen, A.J., 1969. Rip currents: Theoretical investigations. Journal of Geophysical Research 74, 5467–5478.

[4] Branche, C.M., 2001. Lifeguard effectiveness: A report of the working group. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control URL: https://www.cdc.gov/homeandrecreationalsafety/pubs/lifeguardreport-a.pdf.

[5] Brander, R.W., Drozdzewski, D., Dominey-Howes, D., 2014. "Dye in the Water": A visual approach to communicating the rip current hazard. Science Communication 36, 802–810.

[6] Brannstrom, C., Brown, H., Houser, C., Trimble, S., Lavoie, A., 2015. "You can't see them from sitting here": Evaluating beach user understanding of a rip current warning sign. Applied Geography 56, 61–70. doi:10.1016/j.apgeog.2014.10.011.

[7] Brewster, B.C., Gould, R.E., Brander, R.W., 2019. Estimations of rip current rescues and drowning in the United States. Natural Hazards and Earth System Sciences 19, 389–397.

[8] Brighton, B., Sherker, S., Brander, R., Thompson, M., Bradstreet, A., 2013. Rip current related drowning deaths and rescues in Australia 2004-2011. Natural Hazards and Earth System Sciences 13, 1069–1075. doi:10.5194/nhess-13-1069-2013.

[9] Buscombe, D., Carini, R.J., 2019. A Data-Driven Approach to Classifying Wave Breaking in Infrared Imagery. Remote Sensing 11, 859. URL: https://www.mdpi.com/2072-4292/11/7/859, doi:10.3390/rs11070859.

[10] Buscombe, D., Carini, R.J., Harrison, S.R., Chickadel, C.C., Warrick, J.A., 2020. Optical wave gauging using deep neural networks. Coastal Engineering 155, 103593. URL: http://www.sciencedirect.com/science/article/pii/S0378383919301243, doi:10.1016/j.coastaleng.2019.103593.

[11] Buscombe, D., Ritchie, A.C., 2018. Landscape Classification with Deep Neural Networks. Geosciences 8, 244. URL: https://www.mdpi.com/2076-3263/8/7/244, doi:10.3390/geosciences8070244.

[12] Castelle, B., Almar, R., Dorel, M., Lefebvre, J.P., Senechal, N., Anthony, E.J., Laibi, R., Chuchla, R., du Penhoat, Y., 2014. Rip cur-

rents and circulation on a high-energy low-tide-terraced beach (Grand Popo, Benin, West Africa). Journal of Coastal Research 70, 633 – 638. URL: https://doi.org/10.2112/SI70-107.1, doi:10.2112/SI70-107.1.

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

[14] Clark, D., Feddersen, F., Guza, R., 2010. Cross-shore surfzone tracer dispersion in an alongshore current. Journal of Geophysical Research (Oceans) 115. doi:10.1029/2009JC005683.

[15] Clark, D.B., Lenain, L., Feddersen, F., Boss, E., Guza, R., 2014. Aerial imaging of fluorescent dye in the near shore. Journal of Atmospheric and Oceanic Technology 31, 1410–1421.

[16] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 248–255.

[17] Dérian, P., Almar, R., 2017. Wavelet-based optical flow estimation of instant surface currents from shore-based and UAV videos. IEEE Transactions on Geoscience and Remote Sensing 55, 5790–5797.

[18] Ducournau, A., Fablet, R., 2016. Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived SST data, in: 2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS), pp. 1–6. doi:10.1109/PRRS.2016.7867019. iSSN: null.

[19] Dusek, G., Hernandez, D., Willis, M., Brown, J.A., Long, J.W., Porter, D.E., Vance, T.C., 2019. WebCAT: Piloting the development of a web camera coastal observing network for diverse applications. Frontiers in Marine Science 6, 353.

[20] Dusek, G., Seim, H., 2013. A probabilistic rip current forecast model. Journal of Coastal Research 29, 909–925. URL: <GotoISI>://WOS:000321162100015, doi:10.2112/Jcoastres-D-12-00118.1.

[21] Elgar, S., Raubenheimer, B., Guza, R.T., 2001. Current meter performance in the surf zone. Journal of Atmospheric and Oceanic Technology 18, 1735–1746. URL: <GotoISI>://WOS:000171624800010, doi:10.1175/1520-0426(2001)018<1735:cmpits>2.0.co;2. n/a.

[22] Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision 88, 303–338.

[23] da F. Klein, A.H., Santana, G.G., Diehl, F.L., de Menezes, J.T., 2003. Analysis of hazards associated with sea bathing: Results of five years work in oceanic beaches of Santa Catarina state, Southern Brazil. Journal of Coastal Research , 107–116URL: http://www.jstor.org/stable/40928754.

[24] Girshick, R., 2015. Fast R-CNN, in: Proceedings of the IEEE international conference on computer vision, pp. 1440–1448.

[25] Han, J., Zhang, D., Cheng, G., Liu, N., Xu, D., 2018. Advanced deep-learning techniques for salient and category-specific object detection: A survey. IEEE Signal Processing Magazine 35, 84–100.

[26] Han, W., Khorrami, P., Paine, T.L., Ramachandran, P., Babaeizadeh, M., Shi, H., Li, J., Yan, S., Huang, T.S., 2016. Seq-NMS for video object detection. CoRR abs/1602.08465. URL: http://arxiv.org/abs/1602.08465, arXiv:1602.08465.

[27] Holland, K.T., Holman, R.A., Lippmann, T.C., Stanley, J., Plant, N., 1997. Practical use of video imagery in nearshore oceanographic field studies. IEEE Journal of Oceanic Engineering 22, 81–92.

[28] Holman, R.A., Stanley, J., 2007. The history and technical capabilities of Argus. Coastal Engineering 54, 477–491.

[29] Holman, R.A., Symonds, G., Thornton, E.B., Ranasinghe, R., 2006. Rip spacing and persistence on an embayed beach. Journal of Geophysical Research-Oceans 111. URL: <GotoISI>://WOS:000234999800001, doi:ArtnC0100610.1029/2005jc002965.

[30] Horn, B.K.P., Schunck, B.G., 1981. Determining optical flow. Artificial Intelligence 17, 185–203. URL: http://dx.doi.org/10.1016/0004-3702(81)90024-2, doi:10.1016/0004-3702(81)90024-2.

[31] Houser, C., Trimble, S., Brander, R., Brewster, B.C., Dusek, G., Jones, D., 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, 1003–1024. URL: <GotoISI>://WOS:000404797900001, doi:10.5194/nhess-17-1003-2017.

[32] Inch, K., 2014. Surf zone hydrodynamics: Measuring waves and currents. Geomorphological Techniques Chap. 3, Sec 2.3.

[33] Jiang, G.Q., Xu, J., Wei, J., 2018. A Deep Learning Algorithm of Neural Network for the Parameterization of Typhoon-Ocean Feedback in Typhoon Forecast Models. Geophysical Research Letters 45, 3706–3716. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018GL077004, doi:10.1002/2018GL077004.

[34] Johnson, D., Pattiaratchi, C., . Transient rip currents and nearshore circulation on a swell-dominated beach. Journal of Geophysical Research: Oceans 109. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JC001798, doi:10.1029/2003JC001798.

[35] Kang, K., Ouyang, W., Li, H., Wang, X., 2016. Object Detection from Video Tubelets with Convolutional Neural Networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA. pp. 817–825. URL: http://ieeexplore.ieee.org/document/7780464/, doi:10.1109/CVPR.2016.95.

[36] Lima, E., Sun, X., Dong, J., Wang, H., Yang, Y., Liu, L., 2017. Learning and Transferring Convolutional Neural Network Knowledge to Ocean Front Recognition. IEEE Geoscience and Remote Sensing Letters 14, 354–358. URL: http://ieeexplore.ieee.org/document/7829262/, doi:10.1109/LGRS.2016.2643000.

[37] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context, in: European Conference on Computer Vision, Springer. pp. 740–755.

[38] Lippmann, T.C., Holman, R.A., 1989. Quantification of sand bar morphology: A video technique based on wave dissipation. Journal of Geophysical Research: Oceans 94, 995–1011.

[39] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., 2018. Deep learning for generic object detection: A survey. arXiv preprint arXiv:1809.02165 .

[40] Liu, Y., Wu, C.H., 2019. Lifeguarding operational camera kiosk system (LOCKS) for flash rip warning: Development and application. Coastal Engineering 152, 103537.

[41] Luo, J.Y., Irisson, J.O., Graham, B., Guigand, C., Sarafraz, A., Mader, C., Cowen, R.K., 2018. Automated plankton image analysis using convolutional neural networks. Limnology and Oceanography: Methods 16, 814–827. URL: https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lom3.10285, doi:10.1002/lom3.10285.

[42] Lushine, J.B., 1991. A study of rip current drownings and related weather factors. National Weather Digest , 13–19.

[43] Maryan, C., Hoque, M.T., Michael, C., Ioup, E., Abdelguerfi, M., 2019. Machine learning applications in detecting rip channels from images. Applied Soft Computing 78, 84–93.

[44] Nelko, V., Dalrymple, R., 2011. Rip current prediction in Ocean City, Maryland, in: Rip Currents: Beach Safety, Physical Oceanography and Wave Modeling. CRC Press, pp. 45–58. URL: https://www.academia.edu/1017360/Rip_Current_Prediction_at_Ocean_City_Maryland.

[45] Papageorgiou, C.P., Oren, M., Poggio, T., 1998. A general framework for object detection, in: Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society, Washington, DC, USA. pp. 555–562. URL: http://dl.acm.org/citation.cfm?id=938978.939174.

[46] Philip, S., Pang, A., 2016. Detecting and Visualizing Rip Current Using Optical Flow, in: Bertini, E., Elmqvist, N., Wischgoll, T. (Eds.), EuroVis 2016 - Short Papers, The Eurographics Association. p. 115. doi:10.2312/eurovisshort.20161155.

[47] Pitman, S., Gallop, S.L., Haigh, I.D., Mahmoodi, S., Masselink, G., Ranasinghe, R., 2016. Synthetic imagery for the automated detection of rip currents. Journal of coastal research , 912–916.

[48] Pritchard, D.W., Carpenter, J.H., 1960. Measurements of turbulent diffusion in estuarine and inshore waters. International Association of Scientific Hydrology Bulletin 5, 37–50. URL: https://doi.org/10.1080/02626666009493189, doi:10.1080/02626666009493189, arXiv:https://doi.org/10.1080/02626666009493189.

[49] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, real-time object detection, in: The IEEE Con-

ference on Computer Vision and Pattern Recognition (CVPR), p. 0.

[50] Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, pp. 91–99.

[51] Schmidt, W., Woodward, B., Millikan, K., Guza, R., Raubenheimer, B., Elgar, S., 2003. A GPS-tracked surf zone drifter. Journal of Atmospheric and Oceanic Technology 20, 1069–1075.

[52] Stringari, C.E., Harris, D.L., Power, H.E., 2019. A novel machine learning algorithm for tracking remotely sensed waves in the surf zone. Coastal Engineering 147, 149–158. URL: http://www.sciencedirect.com/science/article/pii/S037838391830228X, doi:10.1016/j.coastaleng.2019.02.002.

[53] Viola, P., Jones, M.J., 2004. Robust real-time face detection. International Journal of Computer Vision 57, 137–154. URL: https://doi.org/10.1023/B:VISI.0000013087.49260.fb, doi:10.1023/B:VISI.0000013087.49260.fb.

[54] Zhu, X., Wang, Y., Dai, J., Yuan, L., Wei, Y., 2017. Flow-guided feature aggregation for video object detection, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 408–417.