DEEP LEARNING OBJECT DETECTION APPLICATION TO SURFING WAVE QUALITY

Edward A Atkin[1,2,3] , Jai Davies-Campbell[1] and Rhys McIntosh[1]

Quantitative monitoring is imperative to the sustainable management of coastal resources. Surfing resources have been both created and degenerated or destroyed by activities in the coastal zone. Effective surfing wave quality monitoring requires identification and tracking of the breaking part of the wave and the unbroken wave crest. Remote Camera Systems (RCS) have proven their utility in being able to monitor the coastal zone and provide almost continuous, high frequency data collection. RCSs lend themselves very well to the monitoring of surf breaks which are highly dynamic. The images captured from an RCS monitoring a surf break on the west coast of Aotearoa New Zealand are used to train a Convolutional Neural Network (CNN) to detect the break points (BP), associated crest orientation and relative Still Water Level (SWL) of each instance of breaking waves in each image. Model settings and image annotations were modified over a suite of training cases to improve model efficacy, which was evaluated each epoch of training with mean Average-Precision (mAP; max 1). A mAP of 0.794 was achieved for the BP and Crest Point (CP) CNN, and 8.634 for the SWL. The model was used to detect ~1.6 M objects across ~1 million images, with a mean confidence value of all BP-CP detections of 0.63 and more than 70% of detections being greater than 0.5. This model enables the first automation of meaningful surfing wave quality monitoring.

Keywords: surfing resource, surf break, sustainable management, convolutional neural network, environmental monitoring, remote camera system

INTRODUCTION

Of the metrics for quantitatively evaluating surfing wave quality, which are best summarised in Mead (2000), peel angle (α) (Hutt et al. 2001; Walker et al. 1972; Walker 1974) could be considered the most imperative. Walker et al. (1972) describe α as the angle between the unbroken wave crest and the velocity of the surfer, and it directly relates to the speed at which the breaking part of the wave translates laterally. Open face surfing can be undertaken in a range of breaker types and shapes, from spilling right through to collapsing waves (Battjes 1974; Galvin 1968; Iribarren and Nogales 1949; Mead and Black 2001c; Wiegel 1964). However, if the wave does not peel at a suitable rate (the peel angle is too large or small), at either end of the wave breaking spectrum, the surfer cannot translate laterally along the wave face. Wave shape is a surfing wave quality characteristic that is extremely difficult to monitor. In contrast, peel angle is a surfing quality characteristic that can be readily monitored (Atkin 2010; 2021; Davies-Campbell 2018; McIntosh et al. 2018).

Commonly, the trail of broken white water and the crest of the unbroken part of the wave is used to measure peel angle (e.g., Atkin 2010; Davies-Campbell 2018; Hutt et al. 2001; Mead 2000). Most evaluations of α have made use of aerial photographs (Hutt et al. 2001; Mead 2000; Mead and Black 2001a,b,c; Walker 1974), and more recently repeat satellite imagery (e.g., Atkin et al. 2021). These approaches are often limited both spatially and temporally. Following recommendations in Scarfe (1999), Atkin (2010) used oblique photogrammetry to monitor a Multi-Purpose Reef (MPR) that was constructed at Boscombe in the UK. The approach captured wave breaking conditions during a range of wind, wave, and tidal conditions. The resulting dataset compared well to the MPR's design phase, which showed how the MPR modified α values compared to a proximal control site. The approach by Atkin (2010) was manual collection, georectification, annotation and measurement of peel angles, and, ultimately, was a labour-intensive study.

The use of Remote Camera Systems (RCSs) in the monitoring of coastlines (Holman and Stanley 2007) and their ability to collect repeatable, high frequency data, is ideally suited for studying the dynamic process of wave breaking. The capability to accurately georectify images using Ground Control Points (GCPs) that describe pixel locations in real world coordinates allows for quantitative measurements of geophysical phenomena.

In 2017, a research program focused on addressing the lack of clear quantitative measures and guidelines describing the characteristics and functionality of surf breaks in Aotearoa New Zealand established five RCSs (Atkin et al. 2017; 2019a,b). One of the RCSs was established with a Field of View (FoV) of Waikeri (Manu Bay), on the west coast of Aotearoa New Zealand's North Island (Figure 1). Waikeri is a left-hand (where waves break from left to right when viewed from the land)

[1] eCoast Marine Consulting and Research, 18 Calvert Road, Raglan, 3297, Aotearoa New Zealand. Email: e.atkin@ecoast.co.nz
[2] University of Waikato, Hillcrest, Hamilton, Aotearoa New Zealand
[3] International Association for Surfing Research

point break (Mead 2000), a Surf Break of National Significance (Department of Conservation 2010) and considered world class.

Figure 1. Inset: National location of Waikeri (blue circle). Main: Aerial photo (LINZ) of Waikeri showing camera location (red circle) and the RCS FoV real world footprint (amber dashed line).

During the research project data from the Waikeri RCS was used to develop an automated system to measure peel angle from geo-rectified images (McIntosh et al. 2018). The results were of similar accuracy to manual digitisation; however, the system was time and computationally intensive, and required some filtering of outliers to create the final dataset. Machine learning provides the opportunity to automate tasks and reduce the overhead of manual input. Stringari et al. (2019) used pixel intensity and true colours to develop a one dimensional, cross-shore, broken wave detection algorithm for RCSs in New South Wales, Australia. The work tracked the bore of white water generated during wave breaking as it moved shoreward on a predefined transect, roughly orthogonal to incident wave crests. Kim et al. (2020) used a Neural Network (NN) to track key points of broken and unbroken wave crests in the surf zone. NNs use numerous interconnected nodes, like the synapses of the brain, that process input data. NNs need to be trained to do this with comprehensive datasets and have been shown to be very effective at classifying data and, in particular, the detection of objects in images.

Stringari et al. (2021) also used a pixel intensity approach to isolate foam generated during breaking and distinguish it from passive foam and non-breaking areas. Subsequently, a NN was trained to distinguish between active and passive foam. While this capability was less discrete than the work of Stringari et al. (2019) and Kim et al. (2020), which are spatially limited, these approaches are not definitive enough to describe breaking characteristics for surfing wave quality. A more specific approach is that of Thompson et al. (2021), who used 1,300 labelled images to train a Convolutional Neural Network (CNN) to track the approximate breaking area of the wave, or leading edge of the breaking part of the wave, in oblique images. While the work of Thompson et al. (2021) provides quantitative data to describe surfing amenity, such as ride length, it fails to capture the fundamental components required for measuring α – the unbroken wave crest and point of instantaneous breaking. The work presented here describes how a CNN object detection model (Jocher et al. 2020) was trained to be used in the monitoring of α at Waikeri. The model was selected as it provides accurate object detection faster than other CNNs (e.g., Li et al. 2020), has real-time capability (Malta et al. 2021) and it is efficient (Nepal and Eslamiat 2022 and references therein) - which is important for RCSs with power and compute limitations if the detection model is deployed on site.

METHODOLGY

The following section outlines the steps in the application of a CNN to images of breaking waves collected by the Waikeri RCS and is broken down in to five subsections: Collection, Preparation, Annotation, Training, and Application.

Data Collection

The RCS at Waikeri was established in May 2017 and this work considers images up until June 2020. Ground Control Points (GCPs) for were collected with high accuracy GPS (post processed kinematic) in November 2017. The RCS uses an off-the-shelf security camera with a 9~40 mm, 5 megapixel telephoto lens collecting 1,200 images at 1 Hz every hour, each day during daylight hours. Image cubes refer to a stack of images collected overtime. With the exception of the midday collection, image cubes were processed on site to extract image products (e.g. Bruder and Brodie 2020 and references therein) and subsequently discarded. This work uses the midday image cubes (all 1200 images).

Data Preparation

A calibration of intrinsic characteristics of the camera monitoring Waikeri was undertaken using methodology from Bouguet (2015), the output of which is used in the georectification of camera images and image derived data. The software tools, developed by Bruder and Brodie (2020), and GCPs were used to calculate the position and orientation of the camera to develop a georectification matrix for the conversion of pixel related data into real world coordinates.

To account for any movement in the RCS, images were registered/aligned to permanent features (e.g., buildings; Rodriguez-Padilla *et al.* 2020) relative to a baseline image collected at the time the GCPs were established. The registration was undertaken in three steps: 1) a first pass Discrete Fourier Transform (DFT) method (e.g., Guizar-Sicairos *et al.* 2008) was applied to a cropped subset, greyscale image, focussed on buildings in the Field of View (FoV), and was used to reduce any larger-scale offsets; 2) individual subset DFTs of five notable features in subset images were highlighted using Canny edge detection. If the DFT shift values from a particular subset exceeded three pixels (in x or y), the shift values were omitted; and, 3) the subset DFT shift values were used to estimate a similarity transform, and the transform was used to warp the features to the baseline image (e.g., van der Walt *et al*. 2014).

To reduce image size and focus training, images were cropped at the horizon and in unnecessary land areas, and any remaining areas that were not of interest to training were masked (*e.g.*, car park, foreground building; Figure 2).

Figure 2. Original FoV image from 1300 hrs on the 30th November 2017 (top left), registered image showing slight right to left and up down shift in camera (top right), and grey scale version of the registered image that is cropped and masked (bottom). Note the orange (top row) GCP flag left of centre.

Annotation

Preliminary trials used a mask region-based CNN (He et al. 2017). Using a fixed size box, centred on the leading edge of the breaking part of the wave, as it translates laterally, multiple instances of the broken part of the wave were successfully detected by the neural network. Different sized and shaped boxes were tested, which indicated that model effectiveness was sensitive to the labelling strategy. The labelling of the training dataset in this way, and the resulting identified object (box), provided no means of isolating the crest of the wave, which is integral to the measurement of peel angle.

The data required to train the object detection model (Jocher et al. 2020) consists of an image-normalised x,y location with a width and height to depict a box for each instance of an object, in each image. The features targeted for annotation were the Break Point (BP) and unbroken crest of the wave. The boxes used in the model are orthogonal to the pixel orientation. Wave crest orientation can vary spatially, especially where refraction is taking place, which is common to point breaks (Mead 2000). Wave section length (Moores 2001) was used as a reference for delineating annotation boxes. To accurately annotate these features, an interface was developed to simultaneously view and annotate each image in oblique and geo-rectified views, the latter of which allowed for spatial limitations to be applied (Figure 3).

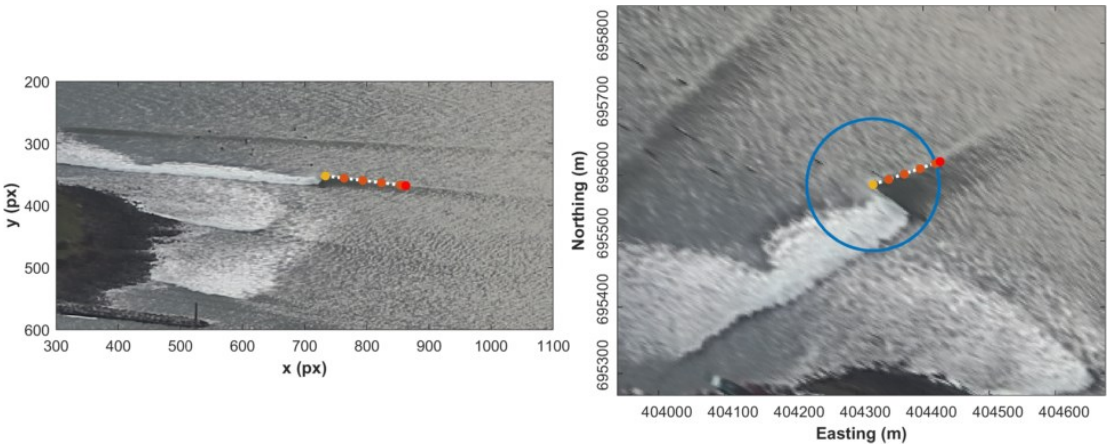

Figure 3. Annotation views of original image (left) and georectified (right), including BP (yellow), crest line (white dashed line) and end point (red), and 5 m CPs (orange).

In ~3,700 images, where each peeling wave was identified: the BP was annotated, and represented with a single x,y location; the BP was used to anchor a 20 m concentric line, within which the crest was considered linear; and, a polyline, anchored to the BP, was aligned to the unbroken crest of the wave and extended beyond the 20 m concentric mark. The output included the coordinates of the BP and the intersect between the polyline and the 20 m circle, the Crest Point (CP). Coordinates were interpolated at 5 m intervals between the BP and the CP to provide CPs at 5, 10, 15, and 20 m from the BP (Figure 3). During annotation, left-hand waves that were not breaking directly on to the shore were digitized. If an image contained no objects, no discernible peeling waves, or if BP were not easily resolved (e.g., during storm events, rain drops on the camera housing window, time of intense sun glare, low light levels etc.) the annotation data for that image was left empty.

Detected object boxes were labelled using the annotation data. Several Training Sets were developed during the study, the key characteristics of which are presented in Table 1. Training Sets 1 and 2 used boxes delimited by the BP and the last CP (*i.e.*, 20 m from the BP); Training Set 2, and all subsequent sets, used greyscale images, which reduced the amount of data that required transferring as well as memory/computational requirements.

From Training Set 3 onwards, the box extent was expanded, doubling in width and height, centred on the BP. This transition was made to include more pixel information in training objects, which was particularly important for far field objects where the difference in the y location of the BP and CP were small, resulting in a narrow object box.

In Training Sets 4 and 5 the second CP (10 m) were used. The third CP (15 m) was used in Training Sets 6, 7, and 8. Training Sets 4 and 6 incorporated an object class which denoted the location of the CP relative to the BP in the oblique image, either above or below, so that crest orientation could be determined in post processing. Of 3,769 annotated breakpoints for the Waikeri training dataset,

98.8% had an BP above the CP. Training Set 8 included only those objects where the CP was below the BP.

Table 1. Training dataset configurations. Colour refers to either Red, Blue, Green (RGB) or Greyscale (GS), Extent refers to the labelling of breaking waves where the Original extended from the Instantaneous Break Point (BP) to the Crest Point (CP), and Expanded quadrupled the area of the bounding box centred on the BP. Class considered the position of the BP relative to the CP within an image.

#	Colour	Extent	CP	CP Class
1	RGB	Original	Last	None
2	GS	Original	Last	None
3	GS	Expanded	Last	None
4	GS	Expanded	Second	Above/Below
5	GS	Expanded	Second	None
6	GS	Expanded	Third	Above/Below
7	GS	Expanded	Third	None
8	GS	Expanded	Third	Below

BP location, at the top of a wave, requires translating to a position at Still Water Level (SWL). Thompson et al., 2021 used a generic offset of 1/5 of the pixel distance between the shoreward edge of the white water bore and the breaking crest to account for the set down Infront of the wave and translate a pixel location to a SWL. Shand et al. (2012), who used a dual camera system to detected breaking waves height, used 1/3 of the position between the crest and trough. Using a constant offset may have been considered effective for waves that are breaking directly toward a cameras FoV However, the assumptions are not applicable to images collected at Waikeri. The broad range of wave breaking types and significant wave refraction occurring at the point break result in Break Points at Still Water Level (BPSWL) being close to the BP right through to being in line with the leading edge of the white water bore. It was observed that not every section of white water bore projects to the base of the wave during breaking, the difference being plunging and spilling waves, which is consistent with Shand et al. (2012).

To account for the BPSWL discrepancy, a second CNN was developed. Annotations for the BPSWL CNN in unrectified images were constrained by the left-right pixel location of previously annotated BPs. The up-down location was estimated using the apparent location of the white water bore and wave breaking shape. Bounding box geometry was a function of the vertical offset in pixels between the annotated BP and BPSWL – the height and width being 2 and 4 times the offset, respectively, the bounding boxes centred on the BP.

Training

The annotated images were split into training, validation, and testing groups using a 7:2:1 ratio. The training stage used data augmentation, which transforms the training data to create a wider range of case images (Jocher et al. 2020). Across some 60 training iterations, of which 38 were usable, as well as the modifications made to the Training Set (Table 2.1), model options and hyperparameters were modified with an aim to improve model accuracy. The key model settings tested in the training phase were model complexity (or size: Small, Medium, or Large), Pre-Trained Weights (PTW), Image Size (ImSz), Batch Size (BchSz), Learning Rate (LR), Weight Decay (WD), and Momentum (M). Training used a freely available, cloud based (Google Colaboratory), high end Graphics Processing Unit (GPU; Tesla K80).

The metric used in this work to determine object detection accuracy is the commonly used mean Average Precision (mAP; Everingham et al. 2015). For each detection, the overlap between the predicted (trained) and ground truth (annotated; validation) is measured by the Intersection over Union (IoU). Precision and recall are calculated based on an IoU threshold; where a prediction is a True Positive (TP) if IoU is equal to or greater than the threshold, otherwise the prediction is False Positive (FP). Precision is the total TP over the total positive results (TP + FP). Recall relies on the number of False Negatives (FN; where the model failed to predict an object) and is the number of TP cases over the sum of total cases (TP + FN). The Average Precision (AP) is the area under the Precision-Recall curve, mAP is the mean of APs per class. mAP@0.5 is the mAP when considering an IoU threshold of 0.5 (50% overlap). Training produces model weights for detection at each epoch and the model weights

associated with the maximum mAP@0.5 for each model run were retained. Training of the BPSWL CNN leveraged the PTW from the BP-CP CNN.

Application

The best performing model weights were used in the detection of objects in ~1 million images of Waikeri taken at midday over a 3-year period. The detection model was implemented locally using a relatively modest GPU, a GTX 1080. For each object detected, the model provided a confidence score based on the probability that the box contained a target object. All detected object coordinates were georectified, with the transformation matrix leveraging the GCPs, translated pixel coordinates and water level; the latter of which was tidally adjusted using a local tide gauge.

RESULTS

Training Success was determined by the maximum mAP@0.5 achieved during training. This results section consists of mAP@0.5 values plotted against model training epoch. The maximum mAP@0.5 value for each training run in a plot is included, with colour matched reference lines also showing the epoch of the maximum.

Figure 4 presents three plots to compare the influence of how Training Sets were prepared (Table 2.1). The model settings for each of the cases are the same. Moving from Colour (RGB) to Grayscale (GS) images results in a 5% increase in max mAP@0.5. Expanding the annotated box (comparing Cases 3 and 4 against Case 5, or Training Sets 1 and 2 against 3, see methods) yielded a 24% increase in mAP@0.5. Incorporating the crest up or crest down class identifier decreased mAP@0.5 by ~20% (Case 10 vs Case 11; Training Set 3 against 4). The difference between Case 5 and Case 10 (same Training Set) was the use of a PTW, which had no discernible effect on the max mAP@0.5 (0.64).

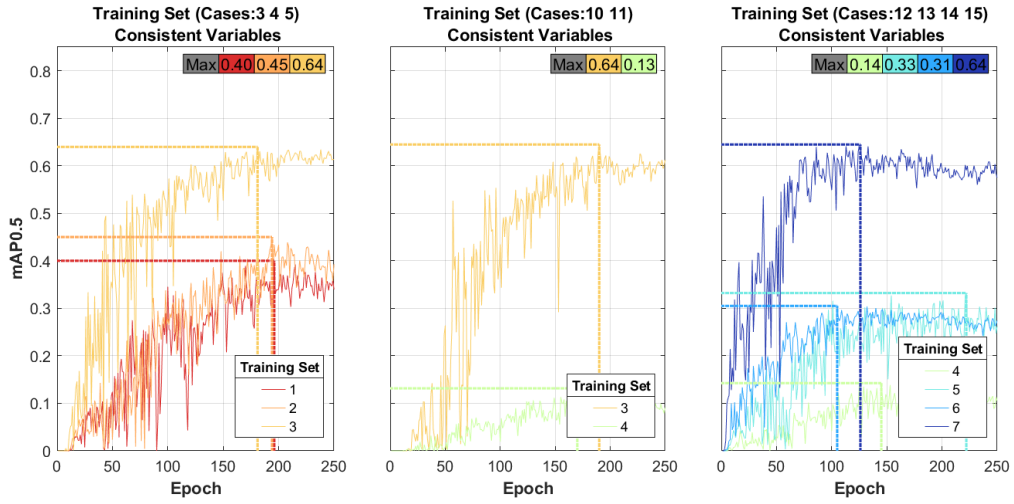

Figure 4. Model training performance plots to compare input training datasets.

Increasing BchSz yields a 1% increase in max mAP@0.5 in two of the three cases (Figure 5). The smaller BchSz results in higher mAP@0.5 values in fewer epochs. Figure 6 illustrates that an image size of 896 pixels (px; in both dimensions) with a small model yields a maximum mAP@0.5 of 0.77, while an image size of 864 px and a medium size model yields a maximum mAP@0.5 of 0.79. The cases used here for model size comparison either vary the PTW, BchSz, or ImSz settings in the model. PTW make little to no difference (1%) to the maximum mAP@0.5 values (Figure 7). Using a BchSz of 80, and a small model, yields the same maximum mAP@0.5 (0.78) as a BchSz of 64 with a larger model (Figure 6). Increasing image size from 640 to 864 increases max mAP@0.5 by 2% (Figure 7).

Maximum mAP@0.5 values have both positive and negative relationships with Learning Rate (LR; Figure 8). Reducing the LR to $1e^{-5}$ results in a very low maximum mAP@0.5, occurring within 250 epochs. The same training parameters were run to 1000 epochs, yielding 0.5 as the maximum mAP@0.5. The LR $1e^{-4}$ largely plateaued before reaching 250 epochs. The difference in LR $1e^{-2}$ and LR $1e^{-4}$ in later cases (29 and 32) was 1%, with the former yielding a higher maximum mAP@0.5. The LR comparison includes variable momentum. Figure 9 shows that decreasing the value of momentum from 0.937 to 0.7 or 0.5 decreases max mAP@0.5 by ~15%; and including the left-right reversals of images in the Training Set decreases max mAP@0.5 by 1%. Using the model's default Weight Decay

(WD) value of $5e^{-4}$ and Training Set 8, yielded a max mAP@0.5 of 0.63. Increasing WD to $5e^{-2}$ decreases the max mAP@0.5 to 0.59. Decreasing WD by 2, 4, and 6 orders of magnitude results in increases of 1%, 4%, and 3 %, respectively.

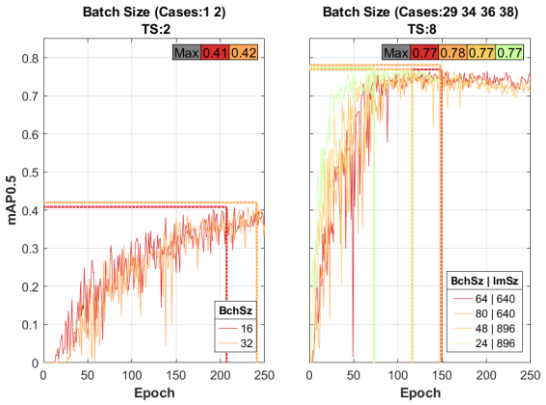

Figure 5. Training performance comparing BchSz.

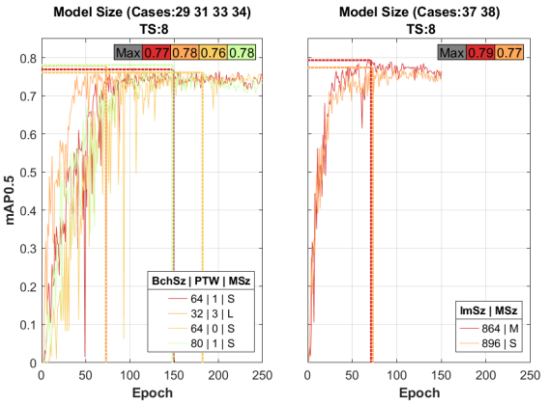

Figure 6. Training performance comparing model size.

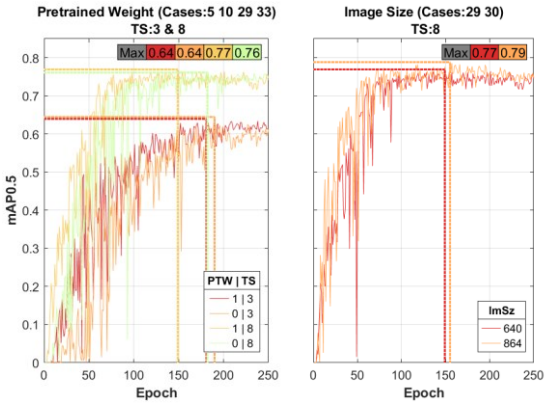

Figure 7. Training performance comparing PTW (left) and image size (right).

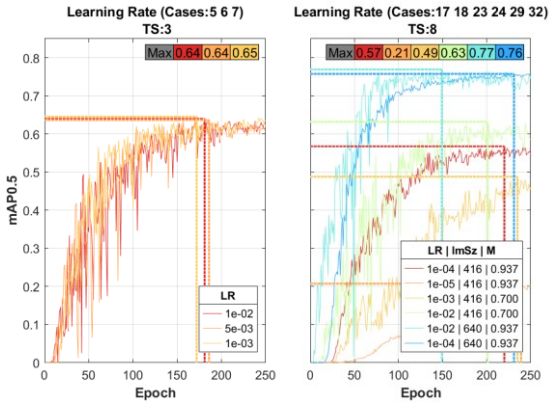

Figure 8. Training performance comparing LR.

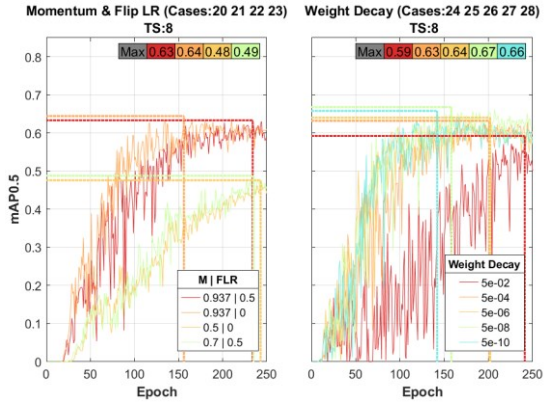

Figure 9. Training performance comparing momentum and FLR (left) and WD (right).

Case 37 yields the highest maximum mAP@0.5 of 0.794 (Figure 6, Figure 10); it uses Training Set 8, default values for LR, M and WD, FLR, a BchSz of 24 (default 16), image size of 864 (default 640) and a medium size model with PTW. Case 30 yielded a max mAP@0.5 of 0.788. The difference between Case 30 and Case 37 is the use of a larger BchSz (64) and a small model. Training of the

BPSWL CNN leveraged the weights from Case 37 with default model settings and achieved a max mAP0.5 of 8.634.

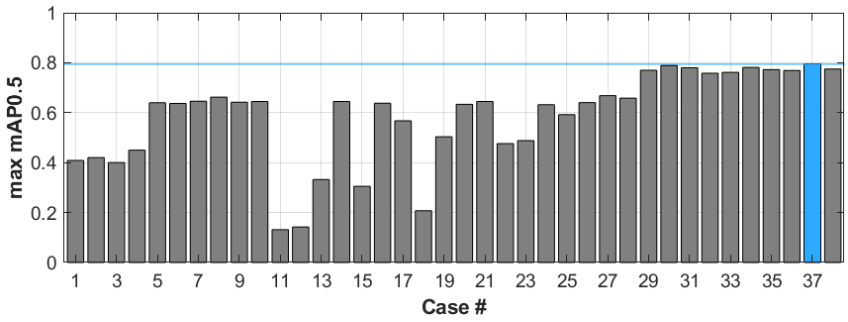

Figure 10. Maximum mAP@0.5 for 38 training cases

Application

The model weights retained from Case 37 were used to detect objects (breaking waves) in ~1 million images. Object detection for a single image occurred in less than 20 milliseconds. Detectable objects were identified in 773,270 of the images, resulting in 1,594,645 break points and crest combinations, each with an associated confidence parameter. Figure 11 demonstrates that confidence is spatially variable and that confidence values range from 0.1 to 0.95. There is a cluster of high confidence values central to the camera's FoV, and an increase in confidence nearer the camera. However, it is not clear if confidence always decreases in the far field as high confidence values are also evident offshore. The boundaries of the FoV appear dominated by low confidence values. The mean confidence value is 0.63, and more than 70% of detections have a confidence value greater than 0.5 (Figure 11).

Peel angle results derived from these detections are presented in detail in Atkin (2021). In summary, the peel angles derived from the automated wave break point and crest detection resulted in average peel angle values that are very comparable to previous studies (Hutt et al. 2001; Scarfe 2008), identified a persistent channel in the seabed and had a peel angle spatial distribution aligned with qualitative descriptions of the surf break (Mead 2000; Scarfe 2008).

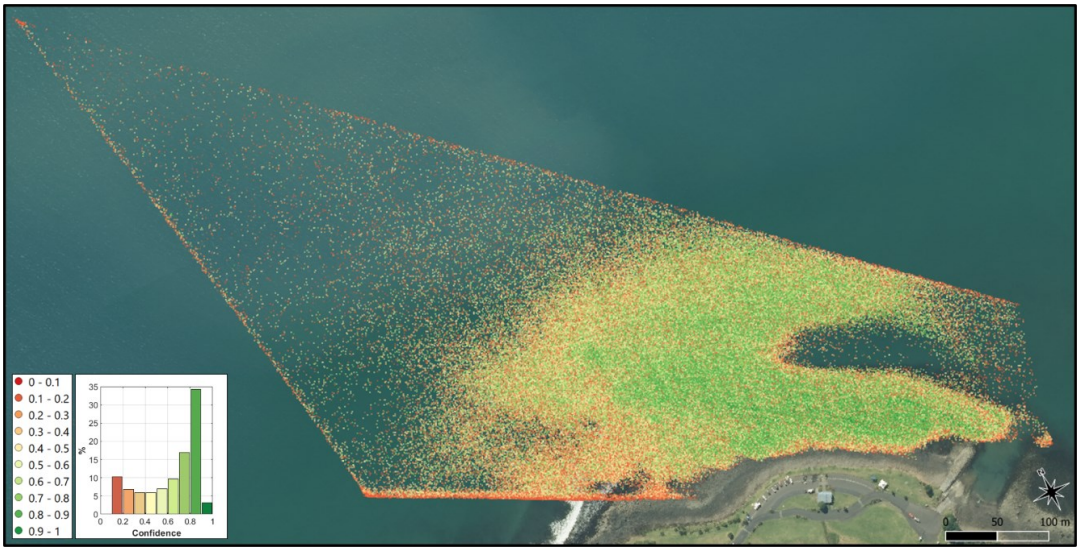

Figure 11. Detected break point confidence. Inset: percentage distribution of confidence values.

DISCUSSION

Of the different model settings and approaches applied to the Training Set in this study, increasing the object detection bounding box had the greatest impact on increasing maximum mAP@0.5. Moving from Colour (RGB) to Grayscale (GS) images provided a 5% increase in max mAP@0.5. These

precision gains are also associated with a reduction in memory allocation, model run time and/or cloud based/network overheads. The improvements in mAP@0.5 made through modification of the Training Set demonstrates the importance of data preparation, which aligns with labelling accuracy being considered the most important component of model performance (e.g., Karimi et al. 2020).

Jocher et al. (2020) indicate that the number of images per class should be no less than 1,500 and that there should be 10,000 or more instances (objects) within the Training Set. Only 3,717 instances were used in this study. Labelling of all instances is also encouraged; however, not all breaking waves were considered surfable, and a clear distinction of what the neural network is trying to achieve, surfability or wave tracking, is required. By virtue of not being able to effectively label the surfing part of the wave, during inclement conditions, a basic surfability threshold or surfing wave quality filter neural network has been developed. In its current state however, the neural network is likely to flag those times where glare from the sun prevented annotation as unsurfable conditions.

Model and image size have also been shown to have a big effect on model training performance. While different model sizes are available, and the results indicate that a larger model will provide higher mAP@0.5 maximums, the choice of model size in this study was limited by the hardware available. A direct model size comparison (all other model settings constant) was not undertaken during the study as the assumption was made that increasing model size would yield higher maximum mAP@0.5 values. The comparison of model size is not direct, either PTW, BchSz or ImSz changes between cases. PTW made little to no difference. A BchSz of 80 and a small model yields the same maximum mAP@0.5 as a BchSz of 64 with a large model, increasing ImSz from 640 px to 864 px increases max mAP@0.5 by 2%. The results provide a good example of the trade-off between model and image size (limited by hardware). BPSWL detection could be improved by using sub-images determined by the BP-CP detection. Using subset images means a higher resolution image could be used in training and/or a larger model.

There is scope for improvement on an essentially well-functioning application of machine learning. While this study was largely focussed on iterating through model settings, Jocher et al. (2020) provide a hyperparameter evolution method, which could be explored in future studies. Improvements to the spatial distribution of confidence values should also be considered. The dominance of low confidence detections on the left-hand side of the RCS FoV, or southern edge of the spatial extent of data in real world coordinates, are attributed to the lack of pixel information to the left of the breakpoint, with training objects including equal extents of the broken and unbroken parts of the wave either side of the Instantaneous Break Point (BP). For confidence in post-processing of detections it may be judicious to exclude points on the FoV boundary, or within a boundary buffer zone related to the width of bounding boxes.

The training set developed in this work is focused on a single surf break. While a wide range of conditions have been captured, a more generic Training Set could be compiled by using not only other RCS systems (e.g., Atkin et al. 2017), but also more commercial surf cameras and/or general images of breaking waves. This could be developed by a wider group of contributors, like the popular baseline training datasets (e.g., Lin et al. 2014), and become effective in a wide range of settings, or develop wave breakpoint and crest specific PTWs. Of relevance to this is the observation that negating the augmentation of vertical mirroring of images (only left-handers) marginally increases max mAP@0.5. The inclusion of both left- and right-hand breaking waves would open up the applicability of this model, especially at more dynamic sites where breaking is less predictable, such as beach breaks.

A restriction included in the method was the omission of any wave where the CP is higher in the image than the BP. This assumption was implemented to bypass the poor performance of the Training Set where a class was included to distinguish relative CP location. The assumption is considered valid, with 98.8% of cases conforming, and excluding the minority (crest above break point) is not a concern in post processing given the frequency of data collection and, therefore, overall points available. However, an approach that may be more holistic is to include a box rotation parameter in the object labelling of the Training Set (e.g., Edeyejedi 2020). The BP could remain central to the object, as the expanded box improved model performance significantly. The crest orientation would be depicted by a box's major axis. The application may also yield better results from bounding circles, over boxes (e.g., Yang et al. 2020). A different model framework may be required for this, and this particular model was chosen for being open source, the quickest object detection model available, and a lightweight framework that could be eventually implemented on low compute RCS.

Being able to quantitatively establish a natural baseline of a coastal resource, such as a surf break, is of great benefit to sustainable management practices. This approach could be readily replicated at other sites; furthermore the application of machine learning for monitoring of coastal hazards could

readily be incorporated into the same system, especially given the speed of detection, which lends itself to real-time monitoring. This application could also be used in live streams of surfing competitions to provide an extra element of analysis, which is encouraged by the rapid detection of multiple objects using a modest GPU.

SUMMARY

The wave breaking characteristic of peel angle is an imperative metric of surfing wave quality and requires data about the path of the broken wave and orientation of the unbroken wave crest. This work considered the application of machine learning techniques to automate the detection, in oblique images collected by a remote camera system, of the instantaneous break point and unbroken crest of waves at Waikeri on the west coast of Aotearoa New Zealand.

While recent work has shown the application of machine learning in surfing amenity, to the authors' knowledge this is the first application of machine learning to detect the fundamental components required to derive peel angle and meaningful surfing wave quality monitoring. While there is scope to improve the model, it has shown rapid and accurate detection capability of an extremely dynamic phenomenon over a high frequency, medium term dataset. This capability has far-reaching potential benefits or the establishment of natural baselines and in the sustainable management of surfing resources.

ACKNOWLEDGMENTS

All work reliant on remote camera systems would never be possible without the generosity of the site hosts, to whom we will always be indebted. The enthusiasm of Amardeep Singh while introducing options for computer vision and object detection, and the ongoing support of Professor Karin Bryan of the University of Waikato are greatly appreciated.

REFERENCES

Atkin, E.A. 2010. *The Impact of an ASR on Breaking Wave Conditions at Boscombe*, UK. Thesis: University of Southampton, UK.

Atkin, E.A. 2021. Machine-learned Peel Angles for Surfing Wave Quality Monitoring. *Proceedings of the 25th Australasian Coasts & Ports 2021 Conference*, Christchurch, New Zealand.

Atkin, E.A., Mead, S.T., Bryan, K., Hume, T. and Waiti, J. 2017. Remote Sensing, Classification and Management Guidelines for Surf Breaks of National and Regional Significance. *Proceedings of the 23rd Australasian Coasts and Ports Conference*, Cairns, Aus., 21-23 June 2017.

Atkin, E.A, Bryan, K., Hume, T., Mead, S. T., and Waiti, J., 2019a. *Management Guidelines for Surfing Resources*. Raglan, Aotearoa New Zealand: Aotearoa New Zealand Association for Surfing Research.

Atkin, E.A, Bryan, K., Mead, S. T., Hume, T., and Waiti, J. 2019b. Management Guidelines for Surfing Resources. *Proceedings of the 24th Australasian Coasts and Ports Conference*, Hobart, Australia, 10-13 September 2019.

Atkin, E.A., Mead, S.T., O'Connell-Milne, S. and Davies-Campbell, J. 2021. *Surf Break of Regional Significance*: Southland. eCoast technical report prepared for Environment Southland.

Battjes, J. A. 1974. Surf Similarity. *Coastal Engineering Proceedings*, 1(14), 26.

Bouguet J.Y. 2015. Camera Calibration. Toolbox for Matlab. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

Bruder, B.L. and Brodie, K.L. 2020. CIRN Quantitative Coastal Imaging Toolbox. *SoftwareX*, 12.

Davies-Campbell, J. 2018. *The Morphology and Surf Conditions of Aramoana Beach, Otago : A Surf Break of National Significance*. Thesis: University of Waikato, Aotearoa New Zealand.

Department of Conservation, 2010. *New Zealand Coastal Policy Statement 2010*. Wellington, New Zealand: Department of Conservation, 30p.

Edeyejedi 2020. Re: *Rotated Bounding Boxes* [Discussion post]. GitHub. https://github.com/ultralytics/yolov5/issues/510

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. 2015. The PASCAL Visual Object Classes Challenge: A Retrospective. *International Journal of Computer Vision*, 111(1), 98-136.

Galvin, C. J., 1968. Breaker Type Classification on Three Laboratory Beaches. *Journal of Geophysical Research,* 73, 3651-3659.

Guizar-Sicairos, M., Thurman, S.T. and Fienup, J.R. 2008. Efficient Subpixel Image Registration Algorithms. *Optics Letters*, 33, 156-158.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. 2017. Mask R-CNN. *In Proceedings of the IEEE International Conference on Computer Vision*.

Holman, R.A and Stanley, J. 2007. The History and Technical Capabilities Of Argus. *Coastal Engineering*, 54, 6, p. 477–491.

Hutt, J. A., Black, K. P. and Mead, S. T. 2001. Classification of Surf Breaks in Relation to Surfing Skill. In: *Black, K.P. (ed.) Natural and Artificial Reefs for Surfing and Coastal Protection, Journal of Coastal Research*, SI No. 29, p. 66-81.

Iribarren, C. R., and Nogales, C. 1949. *Protection des Ports. Section II, Comm. 4, XVIIth Inter. Naval Cong.*, 31-80.

Jocher et al. 2020. *Ultralytics/yolov5.* Version v3.0. DOI: 10.5281/zenodo.3983579

Karimi, D., Dou, H., Warfield, S.K. and Gholipour, A. 2020. Deep Learning with Noisy Labels: Exploring Techniques and Remedies *Medical Image Analysis*, 65.

Kim, J., Kim, J., Kim, T., Huh, D., and Caires, S. 2020. Wave-Tracking in the Surf Zone Using Coastal Video Imagery with Deep Neural Networks. *Atmosphere*, 11(3), 304.

Li, M., Zhang, X., Lei, L., Wang, X. and Guo, X. 2020. Agricultural Greenhouses Detection in High Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD. *Sensors*, 20, 4938.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. 2014. Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision*, p.740-755.

Malta, A., Mendes, M. and Farinha, T. 2021. Augmented Reality Maintenance Assistant Using YOLOv5. *Applied Sciences*, 11, 4758.

McIntosh, R., Atkin, E.A. and Davies-Campbell, J, 2018. Development of an Automated Peel Angle Detection System for the Manu Bay Surf Break. *New Zealand Coastal Society Conference*, Gisborne, 2018.

Mead, S.T. 2000. *Incorporating High-Quality Surfing Breaks into Multi-Purpose Offshore Reefs*. Hamilton, New Zealand: University of Waikato, Ph.D. dissertation.

Mead, S. T. and Black, K. P. 2001a. Field Studies Leading to the Bathymetric Classification of World-Class Surfing Breaks. In: *Black, K.P. (ed.) Natural and Artificial Reefs for Surfing and Coastal Protection, Journal of Coastal Research*, SI 29, pp. 5-21.

Mead, S. T. and Black, K. P. 2001b. Functional Component Configurations Controlling Surfing Wave Quality at World-Class Surfing Breaks. In: *Black, K.P. (ed.) Natural and Artificial Reefs for Surfing and Coastal Protection, Journal of Coastal Research*, SI 29, pp. 22-32.

Mead S. T. and Black, K. P. 2001c. Predicting the Breaking Intensity of Surfing waves. In: *Black, K.P. (ed.) Natural and Artificial Reefs for Surfing and Coastal Protection, Journal of Coastal Research*, SI 29, pp. 51-65.

Moores, A. 2001. *Using Video Images to Quantify Wave Sections and Surfer Parameters*. Hamilton, New Zealand: The University of Waikato, Masters Thesis, 143p.

Nepal, U. and Eslamiat, H. 2022. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors*, 22, 464.

Rodriguez-Padilla, I., Castelle, B., Marieu, V. and Morichon, D. 2020. A Simple and Efficient Image Stabilization Method for Coastal Monitoring Video Systems. *Remote Sensing*, 12, 70.

Scarfe, B.E. 1999. *Hydrography and Photogrammetry: Tools for Artificial Surfing Reef Studies*. MSc Dissertation, University of Otago, New Zealand.

Scarfe, B. 2008. *Oceanographic Considerations for the Management and Protection of Surfing Breaks*. PhD thesis, University of Waikato, New Zealand.

Shand, T.D. Bailey, D.G., and Shand, R.D. 2012. Automated Detection of Breaking Wave Height Technique. *Journal of Coastal Research*, 28(3), 671-682.

Stringari, C.E., Harris, D.L. and Power, H. 2019. A Novel Machine Learning Algorithm for Tracking Remotely Sensed Waves in The Surf Zone. *Coastal Engineering*,147, p.149-158.

Stringari, C.E, Guimarães, P.V, Filipot, J-F., Leckler, F. and Duarte, R. 2021. Deep Neural Networks for Active Wave Breaking Classification. *Scientific Reports*, 11(1), pp.1-12.

Thompson, M.E., Watterson, E. and Baldock, T.E. 2021. Detailed Assessment of Surf Amenity over Reef and Sand Bottom Surf Breaks using Wave Peel Tracking. *Proceedings of the 25th Australasian Coasts & Ports 2021 Conference* – Christchurch, 30 November – 3 December 2021.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T. and the scikit-image contributors 2014. scikit-image: Image processing in Python. *PeerJ*, 2, 453.

Walker, J.R., Palmer, R.Q. and Kukea, J.K. 1972. Recreational Surfing on Hawaiian Reefs. *Proceedings of the 13th Conference on Coastal Engineering*, Vancouver, Canada.

Walker, J.R., 1974, *Recreational Surf Parameters*, Honolulu, Hawaii, USA: University of Hawaii, Look Laboratory Report No. 30, 311p.

Wiegel, R. L., 1964. *Oceanographical Engineering*. N.J: Prentice-Hall.

Yang, H., Deng R., Lu, Y., Z, Zhu, Chen, Y., Roland, J.T., Lu, L., Landman, B.A., Fogo, A.B. and Huo, Y., 2020. CircleNet: Anchor-free Glomerulus Detection with Circle Representation. Medical image computing and computer-assisted intervention, *International Conference on Medical Image Computing and Computer-Assisted Intervention*.