

Letter

A Data-Driven Approach to Classifying Wave Breaking in Infrared Imagery

Daniel Buscombe ^{1,*}  and Roxanne J. Carini ² ¹ School of Earth & Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA² Applied Physics Laboratory, University of Washington, Seattle, WA 98105, USA; rjcarini@uw.edu

* Correspondence: daniel.buscombe@nau.edu

Received: 28 March 2019; Accepted: 05 April 2019; Published: 10 April 2019

Abstract: We apply deep convolutional neural networks (CNNs) to estimate wave breaking type (e.g., non-breaking, spilling, plunging) from close-range monochrome infrared imagery of the surf zone. Image features are extracted using six popular CNN architectures developed for generic image feature extraction. Logistic regression on these features is then used to classify breaker type. The six CNN-based models are compared without and with augmentation, a process that creates larger training datasets using random image transformations. The simplest model performs optimally, achieving average classification accuracies of 89% and 93%, without and with image augmentation respectively. Without augmentation, average classification accuracies vary substantially with CNN model. With augmentation, sensitivity to model choice is minimized. A class activation analysis reveals the relative importance of image features to a given classification. During its passage, the front face and crest of a spilling breaker are more important than the back face. For a plunging breaker, the crest and back face of the wave are most important, which suggests that CNN-based models utilize the distinctive ‘streak’ temperature patterns observed on the back face of plunging breakers for classification.

Keywords: wave breaking; remote sensing; surf zone; machine learning

1. Introduction

In the surf zone, the spatio-temporal patterns and dynamics of wave breaking generate nearshore currents and transport sediment, which changes seafloor topography. This, in turn, affects wave transformation processes, spatial gradients in energy dissipation, and nearshore hydrodynamic circulation patterns [1]. This circulation determines the fate of seabed nutrients, contaminants, and pathogens, and asserts control on the seabed and water column as habitats [2]. Numerical modeling of the onset of wave breaking and the amount of energy lost during breaking (dissipation) usually relies on parameterizations such as Thornton and Guza [3] and Duncan [4]. The evolution of individual waves toward breaking is not fully understood [5], and there is no consensus on a method for predicting the statistics of breaking waves [6]. Therefore, detailed observations of wave breaking are needed to develop improved parameterizations of wave transformation and energy dissipation processes across the surf zone.

Breaking waves are classified into discrete classes, namely collapsing, surging, plunging, and spilling [7]. Plunging (when the wave crest curls forward and abruptly impinges the water surface) and spilling (when an aerated roller cascades down the front face of the wave) are the most common breaker types on open coastlines and sandy beaches. Spilling and plunging breakers have different rates of energy dissipation [8]. Therefore, a robust, fully automated, objective approach to classifying breaker type and other wave properties from remote sensing data would be an extremely valuable tool for studying surf zone energy budgets. Thermal infrared (IR) imagery has been used to study

deep-water wave breaking [9], microscale breaking [10], and surf zone wave breaking [11] because it is relatively insensitive to reflected light from the sun and sensitive to the subtle temperature difference between relatively warm active foam (the foam produced during active breaking) and relatively cool residual foam (the foam left behind in the wake of a breaking wave). Additionally, distinctive streaky temperature patterns have been observed on the back face of plunging waves in IR imagery [12,13]. However, to our knowledge, no attempt has been made to either mechanistically relate the energy content of wave breaking to specific IR temperature patterns, or to harness those IR textures to infer statistical descriptors of spilling and plunging breakers. Here, we successfully attempt the latter.

In recent years, while applied machine learning technologies have developed at an unprecedented pace [14], there has been a concomitant interest in their application to imagery of coastlines for academic research [15–17], beach monitoring, and for leisure purposes. Indeed, the recent proliferation in coastal imaging systems is driven in part by the expectation that such artificially intelligent technologies will make data-driven observations of surf zone hydrodynamics feasible. A specific class of machine learning algorithms called deep convolutional neural networks (CNNs, also known as DCNNs or Convnets), has been shown to produce state-of-the-art performance for a variety of image recognition and classification tasks e.g., [18–21]. A major reputed advantage of deep learning over conventional machine learning is that it does not require manual feature selection and extraction. These practices select or transform image data to make them more amenable to a specific algorithm or to reduce model overfitting, thereby increasing model generality. Another attractive aspect of deep learning models is that they can be used as generic feature extractors by initializing each model “neuron” with a set of weights and biases learned from a different dataset, such as Imagenet [22], a library of millions of labeled generic images. This widespread practice is called ‘transfer learning’ [23] and makes model training considerably faster.

Here, we employ CNN models on a training set of IR images of breaking or near-breaking surf zone waves to create a function that estimates wave breaking type from an arbitrary single image. Feature extraction is automatic, and predictions are made based on image texture related to small-scale spatial patterns in sea surface temperature. The success of the classification is a test of how well generic CNN models, initialized with weights learned from a different dataset and a different set of classes, extract geophysically relevant information from imagery.

2. Materials and Methods

2.1. Field Site and Instrumentation

Observations of breaking waves in the outer surf zone were collected using a thermal IR camera during a field campaign, 7–8 November 2016, at the US Army Corps of Engineers (USACE) Field Research Facility (FRF) in Duck, NC, USA. A DRS UC640-17 long-wavelength (8–14 μm), uncooled VOx Microbolometer IR camera was mounted to the FRF pier and stabilized by four guy lines attached to the pier railings. This camera viewed the sea surface at 45° incidence angle and collected data continuously at 10 Hz. A continuously operating Riegl VZ-400 LIDAR (LIght Detection and Ranging), with ≤ 1 cm accuracy measured the sea surface elevation across a profile intersecting the field of view of the IR camera. An example IR image of a spilling breaker is shown in Figure 1B. For context, Figure 1A shows a surf zone-scale view of the same breaker, and Figure 1C displays the LIDAR profile of the breaker. In this study, 10.5 hours of data were used, during which time wave height and period both varied significantly, from 0 to 5.94 m and 2.32 to 19.36 s, respectively, measured using the LIDAR profiles.

The dataset used here consists of 9400 oblique images. Example IR images of unbroken waves, spilling breakers, and plunging breakers are shown in Figure 2. For each of the training images, breaker type is manually classified. The distribution of pixel intensities in the IR imagery is used to determine whether or not the image contains a breaking wave [11]; then, the image is manually labeled based on the patterns of image intensity. Plunging breakers are typified by an organized streak pattern on

the back face of the wave (Figure 2A), whereas spilling breakers are identified by more unorganized texture on the back face of the wave (Figure 2B).

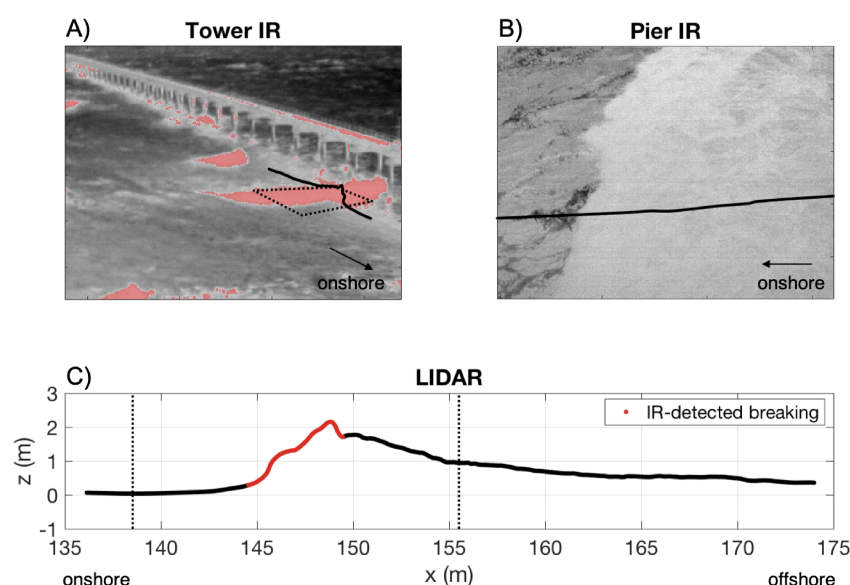


Figure 1. (A) An IR image of the surf zone and the pier are shown with detected breaking indicated by the transparent red layer. The corresponding LIDAR transect (solid black curve) and the pier IR camera field of view (dashed outline) are projected into the image; (B) corresponding example image from the pier IR camera with the LIDAR transect (black line) projected into the image; (C) corresponding example LIDAR transect (same as shown in (A,B)), with active breaking highlighted in red on the front face of the wave, and the field of view of the pier IR camera marked by dashed lines.

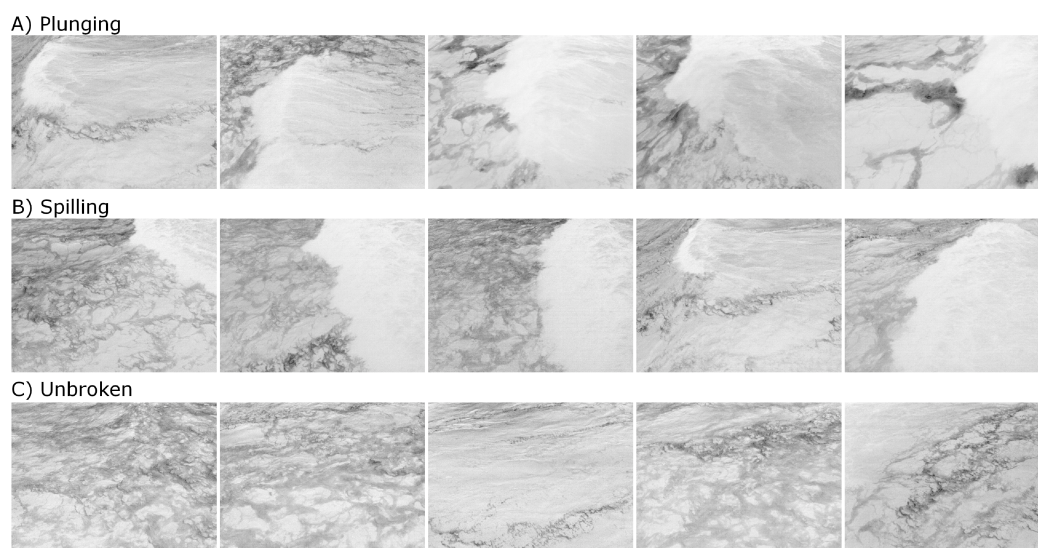


Figure 2. Example IR images of the three categorical wave classes: (A) plunging; (B) spilling; and (C) unbroken waves. The back face of a spilling breaker exhibits a unorganized texture, whereas that of a plunging breaker is characterized by an organized streak pattern.

2.2. Model for Discrete Classification of Breaker Type

A supervised machine learning approach estimates a function f that maps inputs $\mathbf{x}$ (a training set of IR images of waves) to corresponding outputs y (labels of wave breaker class) such that $y = f(\mathbf{x})$. This function is then used to make predictions on unlabeled data. Several open-source CNN architectures have been designed to recognize objects and features in non-specific photographic

imagery [24]. Among numerous suitable, popular, state-of-the-art and open-source frameworks for whole-image classification using CNNs, we choose MobilenetV2 [25], Xception [26], Resnet-50 [27], InceptionV3 [28], Inception-ResnetV2 [29] and VGG19 [19] (in increasing order of number of tunable model parameters). The major difference between Inception-based and Mobilenet-based architectures is that Mobilenet-based models use depthwise separable convolution while Inception V3 uses standard convolutions. In a standard convolution, the filter operates on all image channels of the input image, so the matrix multiplication between the input and filter is multidimensional. However, in a depthwise separable convolution, separate filters operate on each image channel, then output feature maps are generated using a pointwise filter. Inception-ResnetV2 is a hybrid of Inception and Resnet architectures that introduce residual connections that add the output of the convolution operation of the inception module, to the input. The motivating idea behind this is that the next layer will learn the concepts of the previous layer plus the input of that previous layer (the data that was used to learn those concepts). This also allows the model to be much deeper, but with a similar number of model parameters. VGG19 is an older model architecture that is very deep but does not have residual connections or depthwise separable convolutions.

These CNNs automatically generate a hierarchy of features that are learned from input data using a general-purpose procedure. This procedure applies a set of convolution filters to extract local features and spatially shares the parameters of each filter. The weights of the filters in each layer are learned using back-propagation [14], an optimization process that minimizes the error between the output produced by the CNN and the ground-truth label. Each model is implemented using the TensorFlow symbolic math library [30]. Logistic regression on the extracted features is then used for discrete classification. Let $\mathbf{x}$ be the features extracted from the last pooling layer of the CNN.

CNNs are made of many connected layers, and each layer consists of nodes. A node combines input from the input image data with a set of coefficients, called weights, which either amplify or dampen that input. Typically, if a CNN is trained ‘end-to-end’, significance is assigned to inputs with regard to the task that the algorithm is trying to learn (in our case, wave classification) by modifying the values of each weight until optimal classification performance is achieved. Here, we simply take an existing set of weights, determined for each chosen model architecture by a training process for the Imagenet dataset [22]. These input–weight products are summed and then the sum is passed through a node’s activation function; this determines whether and to what extent that signal should progress further through the network to affect the ultimate classification. The set of the final node’s activations is called features, which is used for classification within a multinomial logistic regression. The model so-weighted becomes a generic feature extractor that is optimized for the 1000 Imagenet classes rather than the three wave breaking classes of interest here, but nonetheless we show that the generic CNN is powerful enough to classify those three classes, despite not being optimized. This is an example of so-called ‘transfer learning’ [23].

To handle ambiguity, the most probable wave breaker class, y , is found by transforming $\mathbf{x}$ into a discrete probability distribution over the $k = 3$ possible classes using multinomial logistic regression, known in the field of machine learning as the softmax function [31]. Outputs from this discriminative approach, which models the conditional distribution $p(y|\mathbf{x})$ directly, are found to be consistently superior to the equivalent outputs from a simple generative approach, whereby the posterior label probabilities are found using Bayes’ theorem. The inferior skill of this so-called ‘naive Bayes’ approach is likely due to the extracted image features not contributing independently to the probability of a given classification (i.e., the features are correlated, violating the assumptions of the model).

Models are trained without and with image augmentation. Augmentation is implemented using the following random changes to the images: (1) shifts in either or both image dimensions of up to 10%; (2) rotations up to ± 10 degrees; (3) shear in either axis up to 5 degrees; and (4) zoom up to 20% by image area. Image vertical and horizontal flipping is another common strategy employed to augment training datasets, but it was not employed here because image textures are anisotropic (i.e., the direction of wave propagation is important to image texture). All classified images are randomly split into a

training set and a testing set. The data used to train the CNN models consists of 5455 (7455) images of unbroken waves, 138 (2138) images of plunging breakers, and 1904 (3904) images of spilling breakers without (with) augmentation. The testing data consist of 1820 images of unbroken waves, 47 images of plunging breakers, and 636 images of spilling breakers. The same square area is extracted from each image, and this square region then reduced to 299×299 pixels, for computational efficiency.

3. Results

To assess classification performance, we compare the automated CNN classifications against the test dataset (Table 1). Based on Rank-1 scores, which is the percentage of samples for which the model correctly estimates wave breaker class, there is little difference between MobilenetV2, InceptionV3, and Inception-ResnetV2 (in order of increasing model complexity). Resnet-50 is the worst performing model, without and with image augmentation. The general prediction skill of the model for individual classes is assessed using an F1 score, which is an equal weighting of the recall and precision. Precision and recall are standard accuracy metrics employed when the number of observations belonging to one class (plunging) is significantly lower than those belonging to the other classes (spilling and unbroken), which is the case here. Precision is the proportion of positive identifications that are correct (a precision of 1 means there are no false positives), and recall is the proportion of actual positives identified correctly (a recall of 1 means there are no false negatives). F1 scores (Table 1) reveal greater differences between models and highlight the prediction improvement achieved using image augmentation.

Table 1. CNN classification results. Rank-1 and F1 scores for each of six models without (and with) image augmentation.

Model	Rank-1 Score	F1 Score		
		<i>Unbroken</i>	<i>Plunging</i>	<i>Spilling</i>
MobilenetV2	0.95 (0.94)	0.97 (0.96)	0.85 (0.95)	0.90 (0.89)
Xception	0.94 (0.92)	0.96 (0.95)	0.81 (0.93)	0.88 (0.86)
Resnet-50	0.87 (0.78)	0.92 (0.87)	0.0 (0.6)	0.76 (0.68)
InceptionV3	0.95 (0.94)	0.97 (0.96)	0.69 (0.95)	0.9 (0.9)
Inception-ResnetV2	0.96 (0.93)	0.98 (0.95)	0.77 (0.95)	0.92 (0.89)
VGG19	0.93 (0.90)	0.96 (0.94)	0.24 (0.86)	0.88 (0.83)

For each trained model, a ‘confusion matrix’, which is the matrix of normalized correspondences between true and estimated labels, was used to visualize model skill. A perfect correspondence between true and estimated labels is scored 1.0 along the diagonal elements of the matrix. Misclassifications are readily identified as off-diagonal elements. Systematic misclassifications are recognized as off-diagonal elements with large magnitudes. Comparison of the confusion matrices for models trained without (Figure 3) and with (Figure 4) augmentation shows that the latter ensures individual class predictions have errors of less than 50%, regardless of CNN model used. All augmented models with the exception of Resnet-50 are accurate to within 20% for all classes. Without augmentation, misclassifications are much more likely.

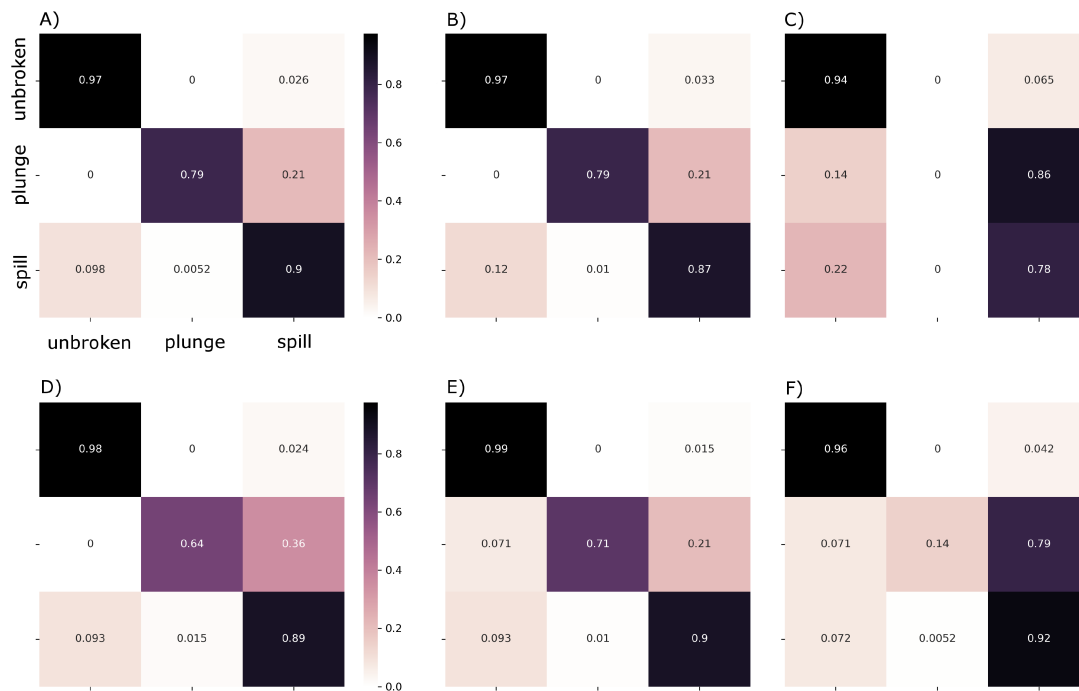


Figure 3. Confusion matrices for models training without image augmentation: (A) MobilenetV2; (B) Xception; (C) Resnet-50; (D) InceptionV3; (E) Inception-ResnetV2; (F) VGG19. For each confusion matrix, the abscissa represents estimated and the ordinate represents true.

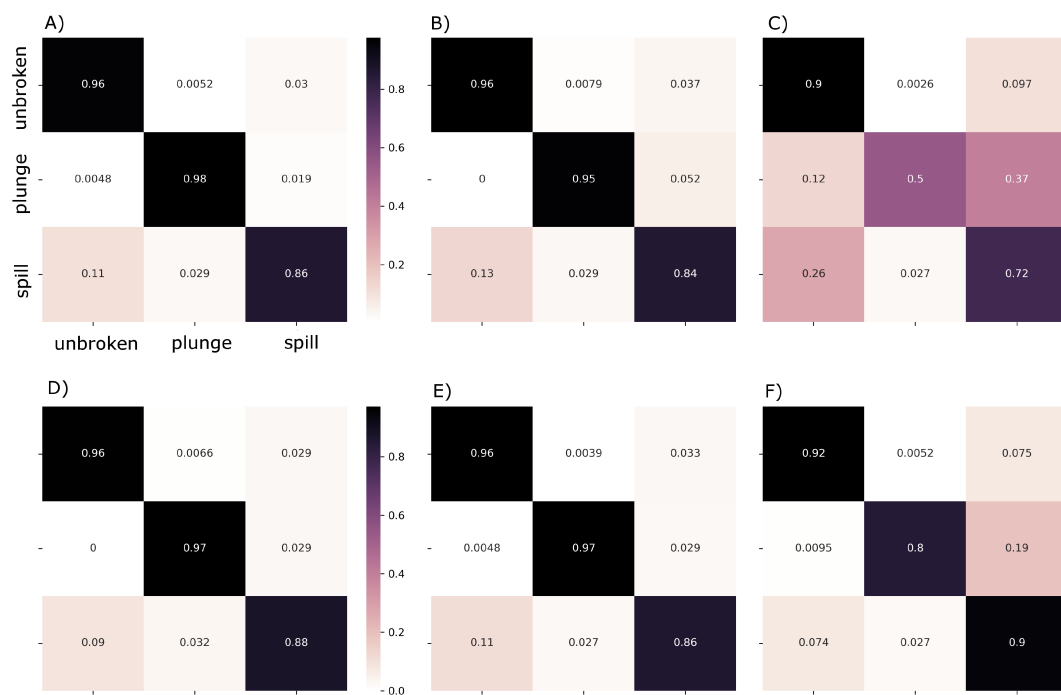


Figure 4. Confusion matrices for models training with image augmentation: (A) MobilenetV2; (B) Xception; (C) Resnet-50; (D) InceptionV3; (E) Inception-ResnetV2; (F) VGG19. For each confusion matrix, the abscissa represents estimated and the ordinate represents true.

4. Discussion

Transfer learning—that is, using a deep neural network with weights learned using another dataset—as a generic feature extractor, combined with a simple logistic regression classifier, was sufficient

for estimating wave breaker class from IR images of the surf zone with high accuracy. For most models, however, image augmentation improved classification performance. This supports the general consensus among machine learning experts that CNNs require large datasets and performance increase is to be expected with more data [32]. This holds even if the classical notion of data information content is challenged by the enormous amount of redundancy in the augmented data. It is widely known that effective use of logistic regression requires large sample sizes [31], hence the importance of augmentation. Augmentation would also further aid end-to-end model training that optimizes the model used as a feature extractor for the specific data set.

The result demonstrates that deep learning is a powerful tool for task-specific classification of dynamic natural features in geophysical imagery. In this specific application, we find that deep learning methods are sensitive to subtle variations of IR image tone, contrast, saturation, and texture that collectively indicate a changing dynamic state. The feature extractors initialized with weights learned on conventional photographic imagery were sufficient to extract the salient features of a geophysical dataset. Unlike many studies applying CNNs to photographic imagery for non-physical purposes, our classification results also suggest that neither residual network connections (such as Resnet-50 and InceptionResnetV2) nor very deep architectures (such as VGG19) are crucial for high accuracy (Table 1). It will be interesting to observe how well these observations hold for other geophysical classification problems.

The global average pooling (GAP) layers in modern CNN feature extractors are crucial to minimize overfitting by reducing the total number of parameters in the model. Furthermore, Zhou et al. [33] and subsequent studies have demonstrated that CNNs with GAP layers that have been trained for a classification task can also be used for object localization. In a conventional sense, this is where an object is in the image. In the present case, we can use this technique to indicate how important each location is with respect to the wave breaker class prediction. The localization is expressed as a ‘class activation map’, where relatively large values indicate regions that are relatively important for the CNN to perform the classification task. We implemented the method of Zhou et al. [33] on five sequential example images of each wave breaker class. Scrutiny of the results in Figure 5 reveals that different features are important, depending on the type of wave or breaker and also on the stage in its temporal evolution. During the passage of a spilling breaker, regions of the image near the wave crest and on the front face of the wave (Figure 5A–E) are most important. For a plunging breaker, (Figure 5F–J), regions of the image near the wave crest and on the back face of the wave are most important. This supports a hypothesis that the CNN is picking up on the distinctive ‘streak’ temperature patterns observed on the back face of plunging breakers [12,13]. Therefore, we argue that not only does the network generally make the correct classification, it also assigns those classes for the right reasons. Such visualization techniques could therefore lead to improved mechanistic understanding of the hydrodynamics or thermodynamics of wave breaking.

Images that were misclassified can almost all be grouped into one of the following four categories: (1) breaking waves just entering the field of view; (2) images containing wake from the previous wave; (3) at the onset of breaking, where a portion of the wave crest is breaking while the rest remains unbroken; and (4) a wave that is in transition from the onset of plunging to the developing breaker stage. Since waves exist somewhere on the continuum between end-member cases of spilling and plunging, those waves whose IR signature is not clearly spilling or plunging pose a challenge to both the DCNN and the eye. Future work to incorporate knowledge of sequential classifications into the CNN algorithm may therefore help to define additional classes, such as the transition between onset and steady state breaking. The ratio of prediction probabilities of spilling and plunging waves may also be a useful metric to determine where a given wave exists on the continuum of breaker type.

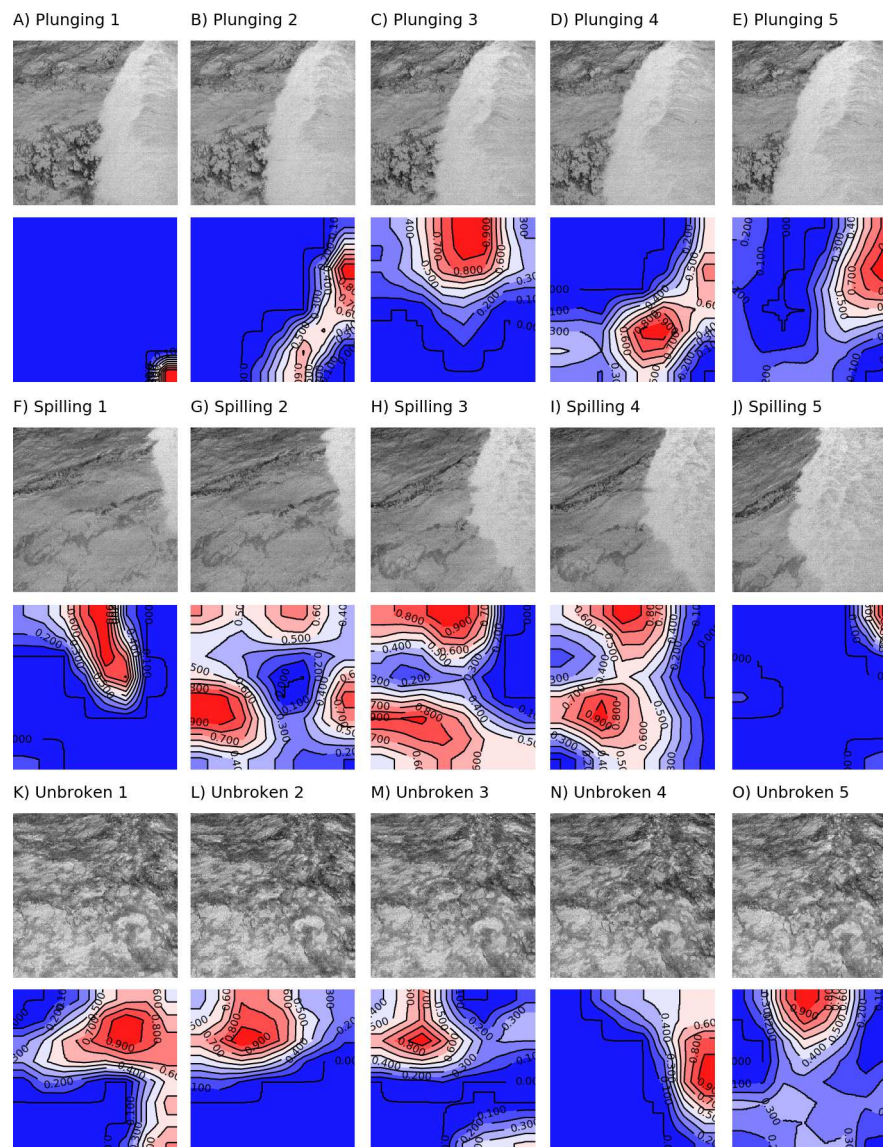


Figure 5. Class activation maps of five sequential example images for each of three wave breaker classes: (A–E) plunging; (F–J) spilling, and (K–O) unbroken. Relatively high values (red colors) are regions of relative importance with respect to the class considered.

5. Conclusions

We successfully classify breaker type in infrared imagery of surf zone waves using deep convolutional neural networks (CNNs). Six CNN-based models are tested using weights learned from generic datasets and a logistic regression classifier. The simplest model (MobilenetV2) performs optimally across all three wave classes (non-breaking, spilling, and plunging), with average classification accuracies of 89% and 93%, without and with image augmentation respectively. Classification error is less than 20% for all classes in all models, except Resnet-50, when training datasets are augmented using random image transformations, and neither residual network connections (Resnet-50 and Inception-ResnetV2) nor very deep networks (VGG19) are required for high classification accuracy.

Class activation maps reveal that the regions of the image that are important for CNN-based class determination correspond to dynamically relevant features of the different breaker types (i.e., the aerated roller on the front face and at the crest of a spilling breaker [3,4], and the strained and streaky back face of a plunging breaker [12,13]). Misclassification is most common for images of ambiguous

breaker type and breakers transitioning from onset to steady state. For these times when a discrete class may not be appropriate to describe the wave state, CNN-based models may be useful for defining the continuum of breaker type and identifying the dynamic features of a developing breaker.

We have presented a technique that may be applied on imagery representing a small spatial footprint of just a few square meters. Therefore, it seems likely that data-driven approaches such as this might be used to identify specific waves in a field of breaking waves by analyzing small regions of imagery with large spatial footprints of thousands of square meters, which would open new research avenues. Another potential next step in application of CNN techniques to IR imagery in the surf zone is image segmentation, which involves classifying every pixel in an image [23,34,35]. Of particular interest would be the spatial extent of active and passive foam, as well as streaks and other thermal patterns, which may be useful for estimating wave energy dissipation.

Author Contributions: D.B. and R.J.C. conceived and designed the experiments; R.J.C. collected the data, D.B. performed the analysis and evaluated the results using validation classifications provided by R.J.C.; D.B. wrote the majority of the paper, while R.J.C. contributed to the introduction and conclusions sections and provided revisions.

Acknowledgments: R.J.C. thanks Kate Brodie, Nick Spore, Jason Pipes, and the entire USACE FRF staff for field support; Dan Clark, Phil Colosimo, and John Mower for IR camera hardware and software development; and Melissa Moulton, Seth Zippel, and Meg Palmsten for fieldwork help. R.J.C. was funded by National Science Foundation Grant No. 1736389 and Office of Naval Research Grant No. N000141010932. This study is fully reproducible using data and code available at https://github.com/dbuscombe-usgs/IR_waveclass.

Conflicts of Interest: The authors declare no conflict of interest.

References

- MacMahan, J.H.; Thornton, E.B.; Reniers, A.J. Rip current review. *Coast. Eng.* **2006**, *53*, 191–208. [CrossRef]
- Defeo, O.; McLachlan, A.; Schoeman, D.S.; Schlacher, T.A.; Dugan, J.; Jones, A.; Lastra, M.; Scapini, F. Threats to sandy beach ecosystems: A review. *Estuar. Coast. Shelf S.* **2009**, *81*, 1–12. [CrossRef]
- Thornton, E.B.; Guza, R. Transformation of wave height distribution. *J. Geophys. Res. Oceans* **1983**, *88*, 5925–5938. [CrossRef]
- Duncan, J. An experimental investigation of breaking waves produced by a towed hydrofoil. *Proc. R. Soc. Lond. A* **1981**, *377*, 331–348. [CrossRef]
- Banner, M.; Peregrine, D. Wave breaking in deep water. *Annu. Rev. Fluid Mech.* **1993**, *25*, 373–397. [CrossRef]
- Filipot, J.F.; Ardhuin, F.; Babanin, A.V. A unified deep-to-shallow water wave-breaking probability parameterization. *J. Geophys. Res. Oceans* **2010**, *115*. [CrossRef]
- Basco, D.R. A qualitative description of wave breaking. *J. Waterw. Port Coast.* **1985**, *111*, 171–188. [CrossRef]
- Iafrati, A. Energy dissipation mechanisms in wave breaking processes: Spilling and highly aerated plunging breaking events. *J. Geophys. Res. Oceans* **2011**, *116*. [CrossRef]
- Jessup, A.; Zappa, C.; Loewen, M.; Hesany, V. Infrared remote sensing of breaking waves. *Nature* **1997**, *385*, 52. [CrossRef]
- Jessup, A.; Zappa, C.J.; Yeh, H. Defining and quantifying microscale wave breaking with infrared imagery. *J. Geophys. Res. Oceans* **1997**, *102*, 23145–23153. [CrossRef]
- Carini, R.J.; Chickadel, C.C.; Jessup, A.T.; Thomson, J. Estimating wave energy dissipation in the surf zone using thermal infrared imagery. *J. Geophys. Res. Oceans* **2015**, *120*, 3937–3957. [CrossRef]
- Watanabe, Y.; Mori, N. Infrared measurements of surface renewal and subsurface vortices in nearshore breaking waves. *J. Geophys. Res. Oceans* **2008**, *113*, C07015. [CrossRef]
- Huang, Z.C.; Hwang, K.S. Measurements of surface thermal structure, kinematics, and turbulence of a large-scale solitary breaking wave using infrared imaging techniques. *Coast. Eng.* **2015**, *96*, 132–147. [CrossRef]
- Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, UK, 2016; Volume 1.
- Holman, R.; Haller, M.C. Remote sensing of the nearshore. *Annu. Rev. Mar. Sci.* **2013**, *5*, 95–113. [CrossRef]
- Baldock, T.E.; Moura, T.; Power, H.E. Video-Based Remote Sensing of Surf Zone Conditions. *IEEE Potentials* **2017**, *36*, 35–41. [CrossRef]

17. Stringari, C.; Harris, D.; Power, H. A novel machine learning algorithm for tracking remotely sensed waves in the surf zone. *Coast. Eng.* **2019**, *147*, 149–158. [[CrossRef](#)]
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
20. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
21. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; Chen, T. Recent advances in convolutional neural networks. *Pattern Recogn.* **2017**, *77*, 354–377. [[CrossRef](#)]
22. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
23. Buscombe, D.; Ritchie, A. Landscape classification with deep neural networks. *Geosciences* **2018**, *8*, 244. [[CrossRef](#)]
24. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
25. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2018**, arXiv:1801.04381.
26. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
29. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
30. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
31. Murphy, K. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, UK, 2012.
32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)]
33. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
34. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
35. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [[CrossRef](#)]

