Coastal Engineering 155 (2020) 103593

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Optical wave gauging using deep neural networks

Daniel Buscombe [a,*], Roxanne J. Carini [b,d], Shawn R. Harrison [c,e], C. Chris Chickadel [b], Jonathan A. Warrick [c]

[a] School of Earth & Sustainability, Northern Arizona University, Flagstaff, AZ, USA
[b] Applied Physics Laboratory, University of Washington, Seattle, WA, USA
[c] U.S. Geological Survey, Santa Cruz, CA, USA
[d] Now at: U.S. Marine Mammal Commission, MD, USA
[e] Now at: Ocean Sciences Division, US Naval Research Laboratory, Stennis Space Center, MA, USA

ARTICLE INFO

Keywords:
Wave monitoring
Remote sensing
Surf zone
Machine learning

ABSTRACT

We develop a remote wave gauging technique to estimate wave height and period from imagery of waves in the surf zone. In this proof-of-concept study, we apply the same framework to three datasets: the first, a set of close-range monochrome infrared (IR) images of individual nearshore waves at Duck, NC, USA; the second, a set of visible (i.e. RGB) band orthomosaics of a larger nearshore area near Santa Cruz, CA, USA; and the third, a set of oblique (unrectified) images from the same site. The network is trained using coincident images and *in situ* wave measurements. The optical wave gauge (OWG) consists of a deep convolutional neural network (CNN) to extract features from imagery — called a 'base model', with additional layers to distill the feature information into lower dimensional spaces, and a final layer of dense neurons to predict continuously varying quantities. Four base models are compared. The OWG is trained for both individual wave height and period, and statistical quantities like significant wave height and peak wave period. The best performing OWG on the IR dataset achieved RMS errors of 0.14 m and 0.41 s for height and period, respectively, capturing up to 98% of the variance in these quantities. The best performing OWG on the visible band rectified dataset achieved RMS errors of 0.08 m and 0.79 s, respectively, for height and period. The same values for the oblique RGB imagery were 0.11 m and 0.81 s for height and period, respectively. Overall, wave height and period accuracy is sensitive to choice of base model; OWGs built upon MobilenetV2 tend to perform worst and those built on Inception-ResnetV2 have the smallest RMS error. The presence or otherwise of residual layers in the model makes little systematic difference to the final OWG accuracy. Smaller batch sizes used in model training tend to result in more accurate OWGs. An out-of-calibration validation, using images associated with wave heights or periods outside the range of values represented in the training data, showed that the ability for OWGs to predict the bottom 5% of low wave heights and the top 5% of high wave heights was reasonably good, but the same was not generally true of wave period. The same framework, not optimized for either dataset, predicts both quantities with high accuracy when trained on imagery, despite the differences in electromagnetic band, perspective, and scale. The OWG estimates wave properties from an image in less than 100 ms on a modestly sized CPU, allowing for the possibility of continuous real-time wave estimates.

1. Introduction

Observation and measurement of wave height and period in the surf zone are important for both monitoring and prediction of nearshore environments. Wave height and period are primary inputs to nearshore wave models that in turn drive circulation and sediment transport predictions, the ultimate goal for management operations and decision making. Routine monitoring for coastal hazards and recreation likewise depend mostly on wave height and period to estimate risks to beachgoers for dangerous surf conditions, such as the prevalence of rip currents. Estimating surf zone wave height with existing techniques is challenging, and most *in situ* wave gauges are located in deep water. Nearshore waves are typically modeled, but less often verified, due to difficulty of deployment and risk of loss in shallow water. Real-time nearshore wave height and period measurements are useful for navigational safety, assessing coastal hazard potential, advising on presence of rip-currents for swimmer safety, surf quality, and water quality.

* Corresponding author.
E-mail address: daniel.buscombe@nau.edu (D. Buscombe).

https://doi.org/10.1016/j.coastaleng.2019.103593
Received 10 March 2019; Received in revised form 16 September 2019; Accepted 2 November 2019
Available online 8 November 2019

Remote sensing of nearshore processes has significant advantages over *in situ* measurements that tend to be limited in spatial and temporal coverage. Visible and thermal infrared (IR) imagery have proven especially useful for capturing spatially extensive observations of hydrodynamics, in particular wave propagation and breaking. Applications of visible band imagery tend to occur at relatively large scales, typically hundreds to thousands of meters in the alongshore (Holman et al., 1993; Holland et al., 1997; Holman and Haller, 2013). When using visible band imagery to study processes associated with individual wave breaking, both reflected light from the sun and residual foam (the foam left behind in the wake of a breaking wave) can individually and collectively overwhelm and obscure the signal of interest, namely the active foam that is generated while a wave is breaking. In IR imagery, however, active foam is differentiable from residual foam and background water, which is one reason IR imagery has been used to study deep-water, microscale, and surf zone wave breaking (Jessup et al., 1997a,b; Carini et al., 2015).

It is possible to infer wave properties from time-series of visible band or infrared imagery. Previous studies have utilized time-series of visible band imagery to extract hydrodynamic properties of nearshore waves and currents, swash and runup (Holman and Guza, 1984; Stockdon et al., 2006; Baldock et al., 2017), breaking wave location and a qualitative measure of breaking intensity (Stockdon and Holman, 2000; Allard et al., 2008), wave period (by monitoring a single pixel in the image over time; Stockdon and Holman (2000)), wave celerity (Stringari et al., 2019), wave dissipation (Aarninkhof and Ruessink, 2004) and attenuation (Pereira et al., 2011), and alongshore currents (Chickadel et al., 2003; Almar et al., 2016). However, these techniques are not always robust to noise (they can be confounded by residual foam or the transition around wave breaking), nor do they always transfer well between sites (due to scale and resolution dependence). Pixel array techniques for computing wave period and celerity require time-series of images, because they rely on tracking features or signals between successive frames. Pixel array techniques are usually sensitive to subjective choices about the position of pixel instruments and the duration over which measurements are made. Further, these methods typically require information about camera geometry to scale and relate the observations to geographical position. No previous generally applicable technique has been proposed and validated to estimate wave height or multiple wave properties from a single image. Stereo imaging (Benetazzo, 2006; De Vries et al., 2011) has the capability of continuously measuring wave height and period using two or more images, but requires camera geometries, significant post-processing, and relies on feature-matching that is computationally demanding and sensitive to image noise. Another alternative is LIDAR (LIght Detection and Ranging) (Irish et al., 2006), using which does not require ground truth to estimate wave properties, but requires significant post-processing. Stationary camera systems have the relative advantage of having no moving parts and can be completely enclosed, often farther away from the sea. It is possible any consumer grade camera with time-lapse capability would provide useful information exploitable by the technique described here.

Deep learning (LeCun et al., 2015; Goodfellow et al., 2016) —a class of machine learning techniques that use large modern neural network models to extract relevant image features automatically —has the potential to be transformative within oceanography. To date, deep learning has been used with remotely sensed imagery to, for example, recognize ocean fronts (Lima et al., 2017), classify coastal environments (Buscombe and Ritchie, 2018), create super-resolution imagery of sea surface temperature (Ducournau and Fablet, 2016), classify plankton (Luo et al., 2018), categorize wave breaking (Buscombe and Carini, 2019) and study internal waves (Pan et al., 2018) and typhoon-induced sea surface temperature cooling (Jiang et al., 2018). These studies demonstrate that deep learning can be a powerful class of tools for analysis of images of dynamic natural features in poly- or monochrome geophysical imagery. For such imagery, solving classification or regression tasks is based upon subtle variations of tone, contrast, saturation, etc., that collectively indicate a different dynamic state. Applications of deep learning in Earth sciences have been reviewed by Reichstein et al. (2019).

Deep convolutional neural networks (CNNs, also known as DCNNs or Convnets) are a specific class of deep learning algorithm that have been shown to produce state-of-the-art performance for a variety of image recognition and classification tasks e.g.(Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Howard et al., 2017; Gu et al., 2017). Conventional machine learning approaches require manual or supervised image feature selection or extraction and sometimes require transformation of the image data so that they are more amenable to a specific algorithm. Deep learning circumvents these practices, which results in increased model generality or decreased overfitting. In conventional machine learning, image band-selection or data dimensionality reduction using ordination techniques are popular (Goodfellow et al., 2016), but for geophysical imagery, such subjectivity built into representations of data involves significant expertise and trial-and-error. Further, variations in lighting, pose, viewpoint, as well as the inherent variation among the features of interest, can make manual or formulaic feature selection and extraction difficult to optimize.

Each layer of a CNN consists of a set of convolution filters connected to the previous and next layers, such that the output of a given filter of a given layer is a function of the outputs of the filters of the previous layer. Through a series of hidden layers, each portion of the image is convolved with a filter set, with each filter designed to search for a particular pattern or feature within the image. CNN-based analysis of geophysical imagery is based upon this hierarchy, which facilitates learning sets of features with different levels of abstraction (Buscombe, 2019). For example, the first few layers identify low-level features such as edges and dark spots. The next few layers then search for medium-level features such as corners, contours, and collections of edges. The final set of layers identify high-level features such as objects and textures with larger structure. This hierarchical design is extremely skillful at recognizing objects or classes in the image, even if they have shifted, shrunk, rotated, or otherwise deformed (He et al., 2016).

Here, the objective is to develop a neural network model framework for generic application, that can predict wave height and period from a given image. The model estimates wave height or period within the image region, rather than for individual waves within the image. Within this model framework, which we call an Optical Wave Gauge or OWG, feature extraction is automatic, and predictions are made on image textures that relate to wave geometry and —in the case of the IR imagery —also small-scale spatial patterns in sea surface temperature. We test the OWG with three image datasets: one consisting of short-range oblique (unrectified) IR imagery of individual breaking and unbroken waves; orthomosaics of rectified visible-band imagery of a larger nearshore area; and finally oblique (unrectified) visible-band images from the same area.

2. Field sites and data

2.1. Close-range infrared imagery and wave measurements

Close-range thermal IR images of breaking waves in the surf zone (Fig. 1) were collected during a field campaign, 7–8 November 2016, at the US Army Corps of Engineers (USACE) Field Research Facility (FRF) in Duck, North Carolina, United States. A DRS UC640-17 long-wavelength (8-14 μm), uncooled VOx Microbolometer IR camera was mounted to a small tower secured to the FRF pier and viewed the sea surface at 45° incidence angle, which resulted in a 20 m wide field of view. The camera collected images continuously at 10 Hz. During the 10.5 h of data collected over 48 h used for this study, individual wave heights and periods varied significantly, from 0 to 5.94 m and 2.32 to 19.36 s, respectively. These values represent the full range of wave heights and periods measured. There was a storm offshore on

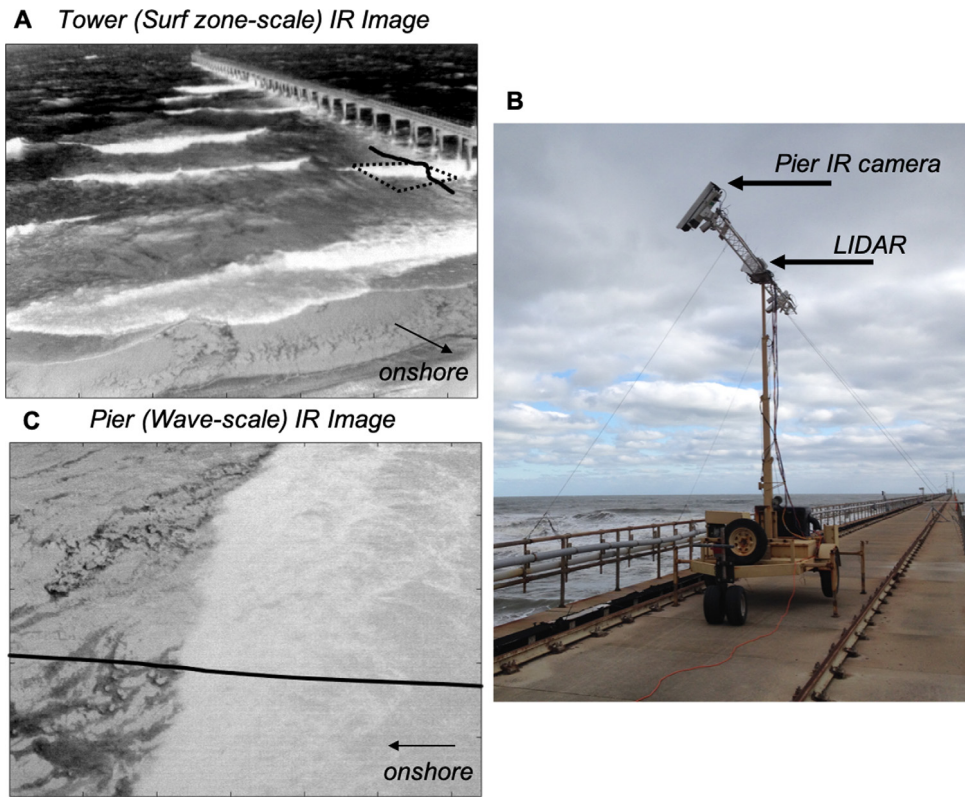

Fig. 1. (A) Infrared (IR) image of a large nearshore area from an IR camera mounted to a high tower. Also indicated is the field-of view of a second IR camera (B) and 1D transect scanned by a LiDAR (B), both overlooking the surf zone from a pier. The imagery used in this study was the smaller footprint wave-scale imagery (C) from the pier-mounted IR camera. The dataset exemplified in (A) was not used in the present study.

the 7th that never made landfall, but was responsible for the longer period, larger wave height swell that arrived on the 8th. Wave direction varied slowly. Significant wave heights, peak periods, direction, and tidal elevations during the field campaign are shown in Fig. 2.

Wave heights and periods were measured with ≤ 1 cm accuracy by a Riegl VZ-400 LIDAR scanning continuously along a sea surface profile intersecting the field of view of the IR camera. The data consist of 9400 oblique images (Fig. 3) with associated wave height and period measured by the LIDAR. The IR imagery (in 8-bit grayscale format) were cropped from 640 × 480 to 480 × 480 then downsized to 128 × 128 pixels for more efficient model training —an image size that can be accommodated in training and validation batches of up to 128 images, with the largest model, on a Graphics Processing Unit (GPU) with 8 GB of memory.

2.2. Visible-band imagery and wave measurements

The data consist of 980 images of the nearshore of Sunset State Beach, Watsonville, California (approximately 15 miles southeast of Santa Cruz), United States, between 6 December 2017 and 30 January 2018, with associated wave height and period measured using an instrumented tripod in 12–14 m of water depth immediately offshore of the site. Significant wave height H_s and peak wave period T_p time-series come from a Nortek Signature 1000 acoustic Doppler current meter. During the 2 months of data used for this study, significant wave height and peak wave period varied from 0.39 to 2.56 m and 7 to 23 s, respectively. Unlike for the IR dataset described above, where the field of view is small but sample frequency is sufficiently high to image every wave that moved through the field of view, the visible-band imagery captures a larger area of the surfzone but at a slower sample frequency (every 30 min). For these data, estimating the height and period of individual waves is not the goal. Rather, waves were measured by 20-minute ADCP bursts every hour, and so not every image corresponds

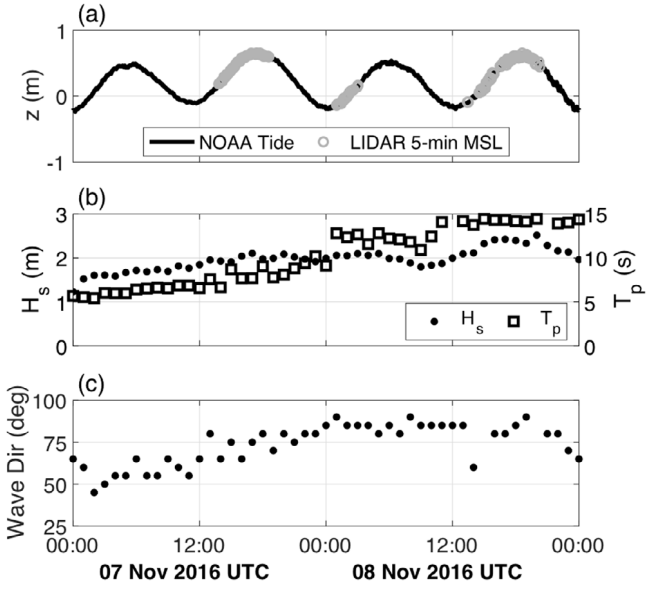

Fig. 2. Tide and wave conditions during 7–8 November 2016, at the US Army Corps of Engineers (USACE) Field Research Facility (FRF) in Duck, North Carolina: (A) Tidal elevation (black line), and the 10.5 total hours of data collection used for this study (light gray circles); (B) Significant wave height H_s and peak wave period T_p, and (C) wave direction.

to a period during which an ADCP burst was collecting. The bulk wave statistics that are the target prediction of the model were interpolated over the timing of the imagery.

The imagery was created by a long-term 2-camera Argus (Holman and Stanley, 2007) station (Fig. 4A) used for remotely sensing coastal change. The two oblique camera views (Fig. 4B, C) were rectified onto

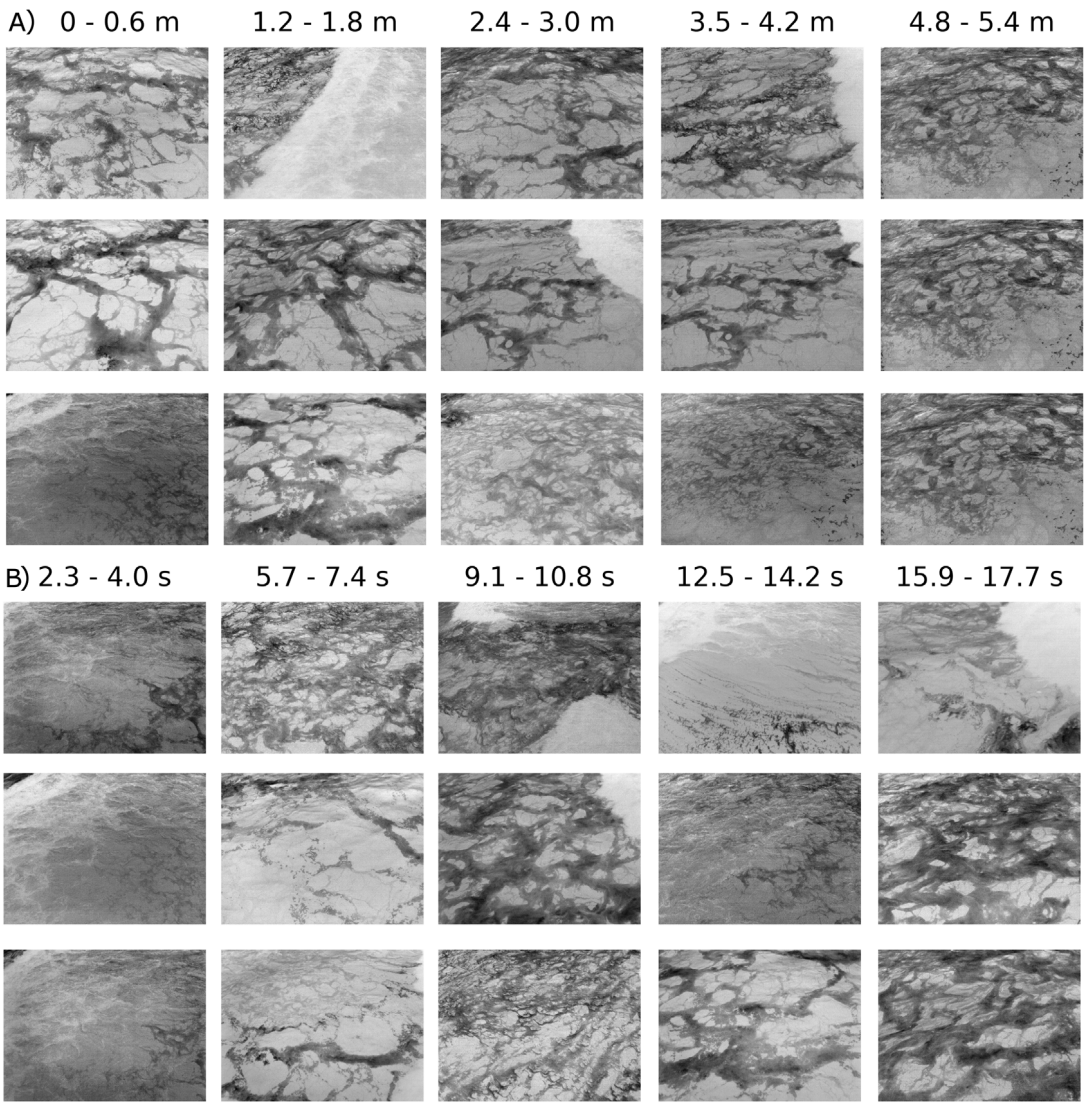

Fig. 3. Example randomly selected IR images associated with increasing individual wave height (A) and period (B). The three rows in A and the three rows in B depict three random samples. In IR imagery, light regions are relatively warm and dark regions are relatively cool.

the horizontal plane of the tidal water level, and merged together into a single planar view. Two datasets have been derived for training OWGs, namely 'orthomosaic' (or 'rectified') and 'oblique' (or 'unrectified') imagery.

The orthomosaics (Fig. 4D) have a rectified horizontal pixel footprint of 0.5 × 0.5 m covering a region of 1.1 × 1.1 km. The station takes 10-minutes of video at 2 frames per second every 30 min of daylight hours, from an oblique vantage point nearly 63 m above mean sea level atop an adjacent coastal bluff. Original images were 2201 × 1901 × 3 pixels in 8-bit RGB format, but were cropped to square, then downsized to 128 × 128 × 1 pixels in 8-bit grayscale format for more efficient model training on a GPU. Each pixel in the downsized imagery is 7.43 × 7.43 m. The merging of the two camera views and differences in the camera color balance between the two cameras causes diagonal seam lines in several of the images (Fig. 5).

The oblique imagery consists of the imagery from camera two (Fig. 4) with the farthest field-of-view alongshore (Fig. 4C, Fig. 6). Original images were 2448 × 2048 × 3 pixels in 8-bit RGB format, but were cropped to square, then downsized to 128 × 128 × 1 pixels in 8-bit grayscale format. Image quality of neither dataset was considered in the model training, therefore relatively rare images with sun glint, dirty lenses, or rain-drops on the lens were not removed and potentially impacted the results negatively.

3. Methods

In order to estimate wave height or period from input imagery, we create a generic CNN architecture (Fig. 7) based on a core feature extractor, called a base model, with (1) a batch normalization layer before and after the base model, which applies a transformation that maintains the mean neuron activation close to 0 and the activation standard deviation close to 1 (Ioffe and Szegedy, 2015) and the result of which is fed into (2) a 2D global average pooling (GAP) layer that averages the activations across each part of the image, then (3) a dropout layer to avoid overfitting (Srivastava et al., 2014), with a dropout rate of 0.5, and finally (4) a dense predicting layer with no activation. The first batch normalization is applied to the raw image, and the second to the activation maps from the base model. GAP layers are used to reduce the spatial dimensions of each of the three-dimensional tensors associated with each pixel of the input image, from h×w×d to 1 × 1×d, by averaging over h and w. This has the effect of reducing the total number of parameters in the model, thereby minimizing overfitting.

Four base CNN models were compared, with a range of architectures and sizes (Table 1): MobilenetV1 (Howard et al., 2017), MobilenetV2 (Sandler et al., 2018), InceptionV3 (Szegedy et al., 2016), and Inception-ResnetV2 (Szegedy et al., 2017). The major difference

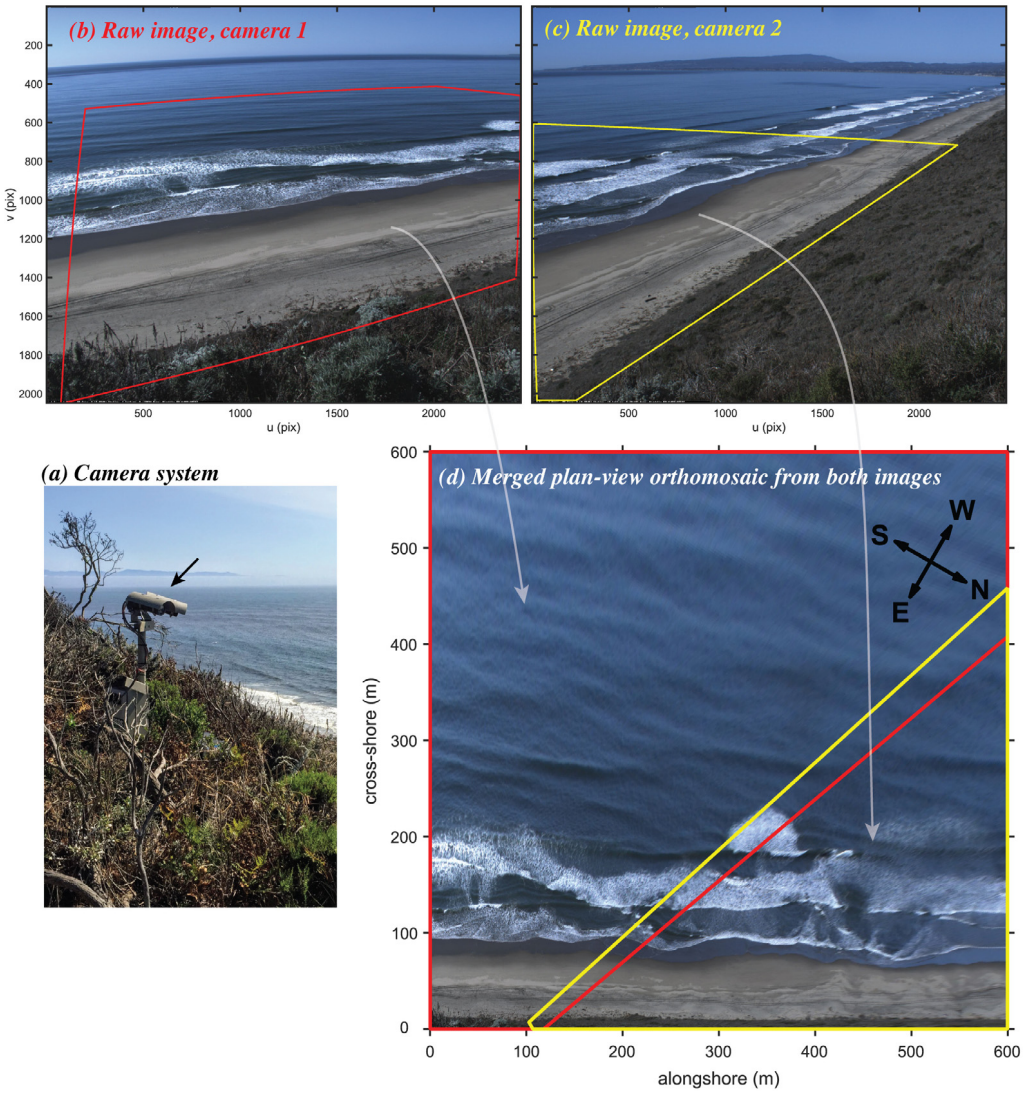

Fig. 4. (A) Two camera coastal monitoring system; (B) and (C) example snapshot images of the left and right camera; and (D) the orthomosaic of the two rectified images.

between the MobilenetV1 and MobilenetV2 models, and between the InceptionV3 and Inception-ResnetV2 models is the presence or absence of residual layers in the model architecture. A hidden layer in MobilenetV1 or InceptionV3 learns to calculate $y = f(\mathbf{x})$ (y and x are generic outputs and inputs), whereas a residual neural network hidden layer in the MobilenetV2 or Inception-ResnetV2 models calculates $y = f(\mathbf{x}) + \mathbf{x}$ (He et al., 2016; Chollet, 2017). In other words, data are allowed to flow through both non-linear activation functions (in $f(\mathbf{x})$) as well as through the network directly ($+\mathbf{x}$). 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 (or even smaller, as in the case of MobilenetV2) number of model parameters.

Each OWG was retrained end-to-end, which means it was initialized with random numbers for neuron weights and biases, then during training the value of those parameters was optimized by minimizing the discrepancy between known and estimated wave height or period. Each OWG was trained with different batch sizes (16, 32, 64, and 128 randomly selected pairs of images and labels) in order to examine their relative effects on results. Sampling used stratification, whereby it was equally likely with the large number of trials to select imagery corresponding to ten equally spaced categories of wave height or period. This was designed to avoid introducing any bias associated with selecting wave height/period magnitudes based on their relative proportion within the training image set, which would tend to preferentially select mid-sized waves over extreme values.

The ideal batch size is one that is small enough to create a regularizing effect on the network, resulting in lower generalization error, but large enough that each training epoch is subject to enough example images to update weights and biases whose values then fluctuate less upon successive epochs, thereby increasing model stability. One training epoch means that the learning algorithm has made one pass through the training dataset, where examples were separated into randomly selected batches of images. Models typically trained for between 100 and 200 epochs before the criterion was met to stop training early. The number of training steps per epoch was computed as the number of training images divided by the batch size. Upon each step, the gradients of the network are updated and new weights assigned to each neuron. Each of the resulting 96 OWGs, consisting of 32 OWGs (16 for wave height and 16 for wave period) for each of the three datasets, were trained for a maximum of 200 epochs. Models stopped training early (i.e. before 200 epochs) if the validation loss failed to improve for 15 consecutive epochs.

The loss function we used during training was mean squared error (the mean squared error of predicted wave height or period compared to observed). Training utilized the popular Adam algorithm (Kingma and Ba, 2014) for stochastic optimization, with parameters β_1=0.9

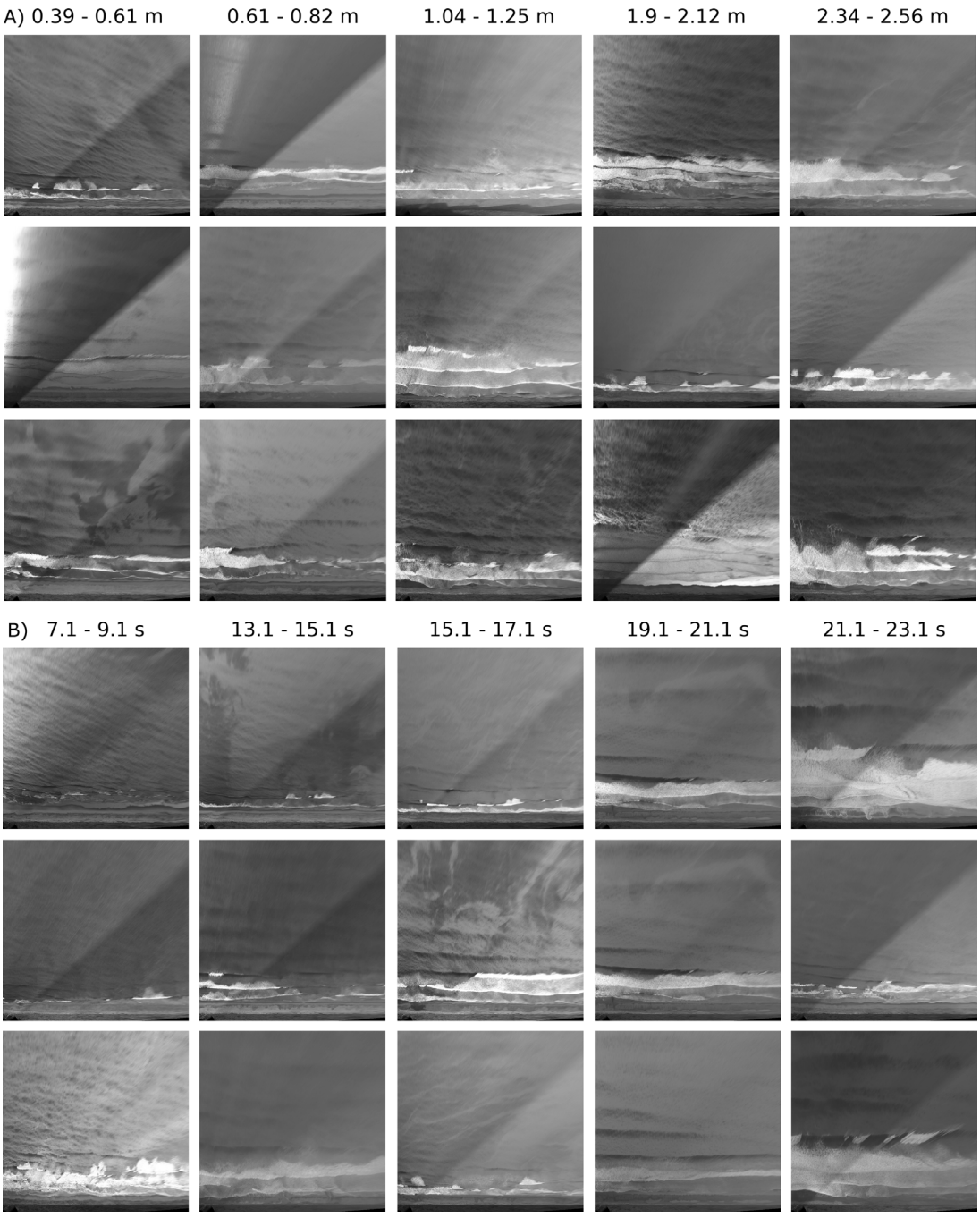

Fig. 5. Example nearshore images associated with increasing H_s (A) and T_p (B). The three rows in A and the three rows in B depict three random samples. Textures associated with larger wave heights show more energetic wave breaking, and larger shadows on the front faces of waves. Image textures associated with long period waves show more organized and regular crests and more energetic wave breaking, and potentially wider surf zone and further offshore onset of wave breaking. The wave gauge was located at the seaward edge of the imagery.

and β_2=0.999. During training, the learning rate was automatically reduced when the loss function stabilized, i.e. when its value stopped decreasing. The learning rate was reduced by a factor of 0.8 after 10 epochs had elapsed with no improvement. A lower bound on the learning rate was set at 0.0001. All OWGs were trained with image augmentation, implemented using random: (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. The general consensus among machine learning experts is that incorporating more data will increase CNNs performance (LeCun et al., 2015), even if the enormous amount of redundancy in the augmented data defies the classical notion of data information content. For the visible band datasets, augmentation resulted in 3000 training and 1000 validation images generated from the original 980. Out of the original 9400 IR images, augmentation resulted in 13,400 training and 6600 validation images.

We did not include the images associated with the top 5% and bottom 5% of wave height or period values in model training, so we could independently test how well the model predicts outside of the range of values used to train it. We refer to this as the 'out-of-calibration' validation. Specifying too few steps per epoch can cause out-of-calibration errors to become large. Of the remaining 90% of images representing 90% of measured wave heights or periods, 60% (54% of all data) were used to train each model, and the remaining 40%

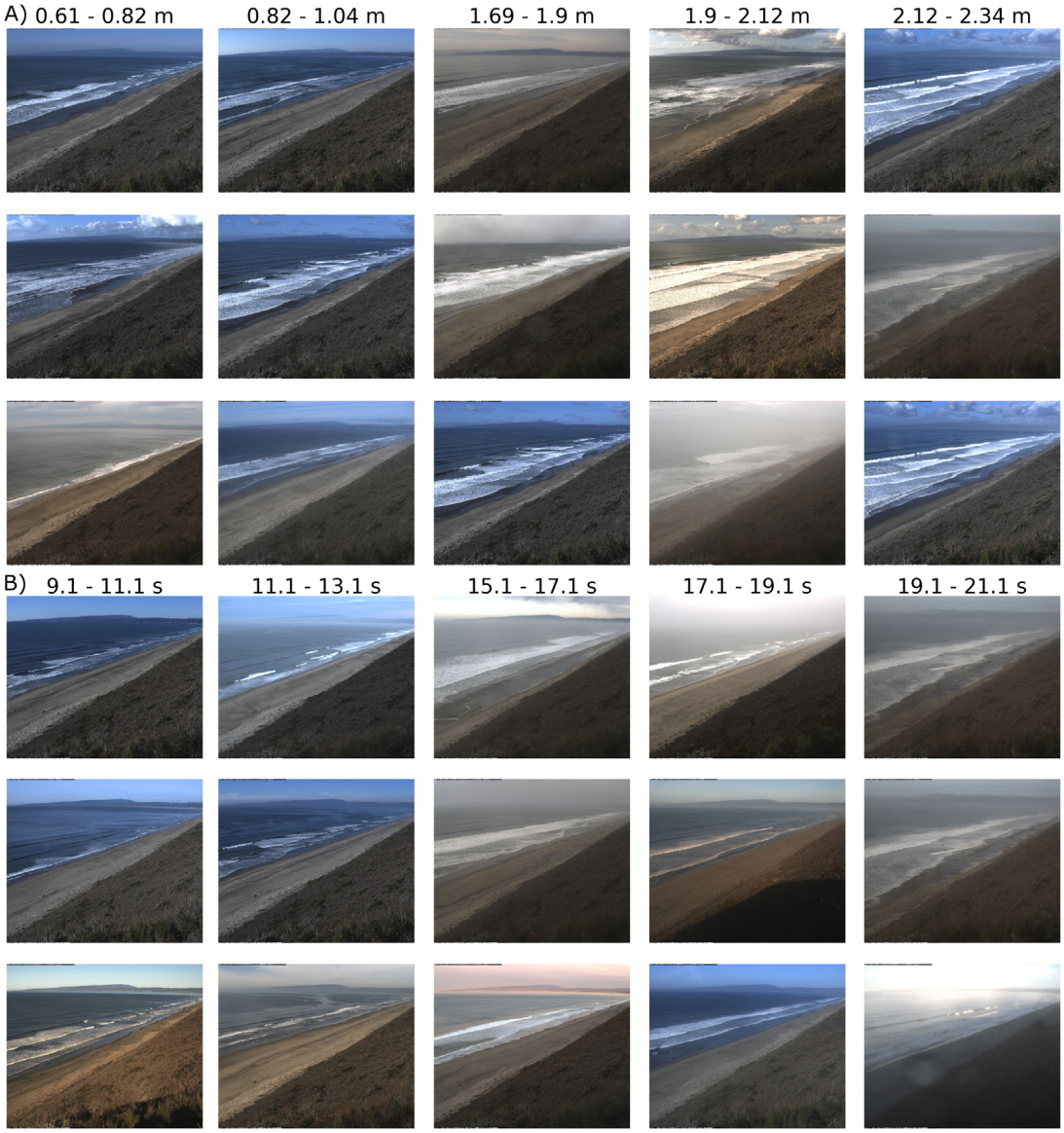

Fig. 6. Example nearshore images associated with increasing H_s (A) and T_p (B). The three rows in A and the three rows in B depict three random samples. Textures associated with larger wave heights show a wider surf zone and further offshore onset of wave breaking. Image textures associated with different period waves show variation in number and spacing of the breakpoint.

(36% of all data) were used to validate each model.All models were trained using the Tensorflow backend to the keras (Chollet et al., 2015) python module, on a 8 GB GeForce RTX 2070 with 2304 CUDA cores. The theoretical resolving power of a network might be thought of as a quantization problem: the resolution of the measurement is the range of the variable in the training data divided by the number of neurons in the final dense layer (Table 1), which in this case is O(0.01) m and O(0.01) s, for wave height and period respectively. Because images are randomly selected during training, the results from the data-driven models presented here are not affected by serial dependence (such as how successive 20 m-footprint images could contain parts of the same wave).

4. Results

The best performing OWG on the IR dataset achieved RMS errors of 0.14 (0.08) m and 0.41 (1.65) s (values in parentheses are for out-of-calibration samples), for height (Table 2) and period (Table 3) respectively, capturing up to 98% of the variance in these quantities. The best performing OWG on the visible band rectified dataset achieved RMS errors of 0.08 (0.14) m and 0.79 (3.44) s for height (Table 2) and period (Table 3), respectively. The same values for the oblique RGB imagery were 0.11 (0.18) m and 0.81 (1.37) s for height and period, respectively.

Overall, wave height and period accuracy is sensitive to choice of base model. OWGs built upon MobilenetV2 tend to perform worst, whereas OWGs built on Inception-ResnetV2 tend to have the smallest RMS error. Using Inception-ResnetV2 as a base model, mean RMS error for wave height were 17, 11, and 14 cm, for the oblique IR, orthomosaic RGB, and oblique RGB datasets, respectively, compared to 60, 26, and 24 cm for OWGs based on MobilenetV2 (Table 2). For wave period, Inception-ResnetV2-based OWGs had mean RMS errors of 0.53, 0.98, and 1.01 s respectively for the three datasets, compared to 0.54, 3.15, and 3.04 s for OWGs based on MobilenetV2 (Table 3).

The presence or otherwise of residual layers in the model makes little systematic difference to the final OWG accuracy (Tables 2 and 3); RMS errors of MobilenetV1 (without residual layers) tend to be smaller than those of MobilenetV2 (with residual layers), and only the best

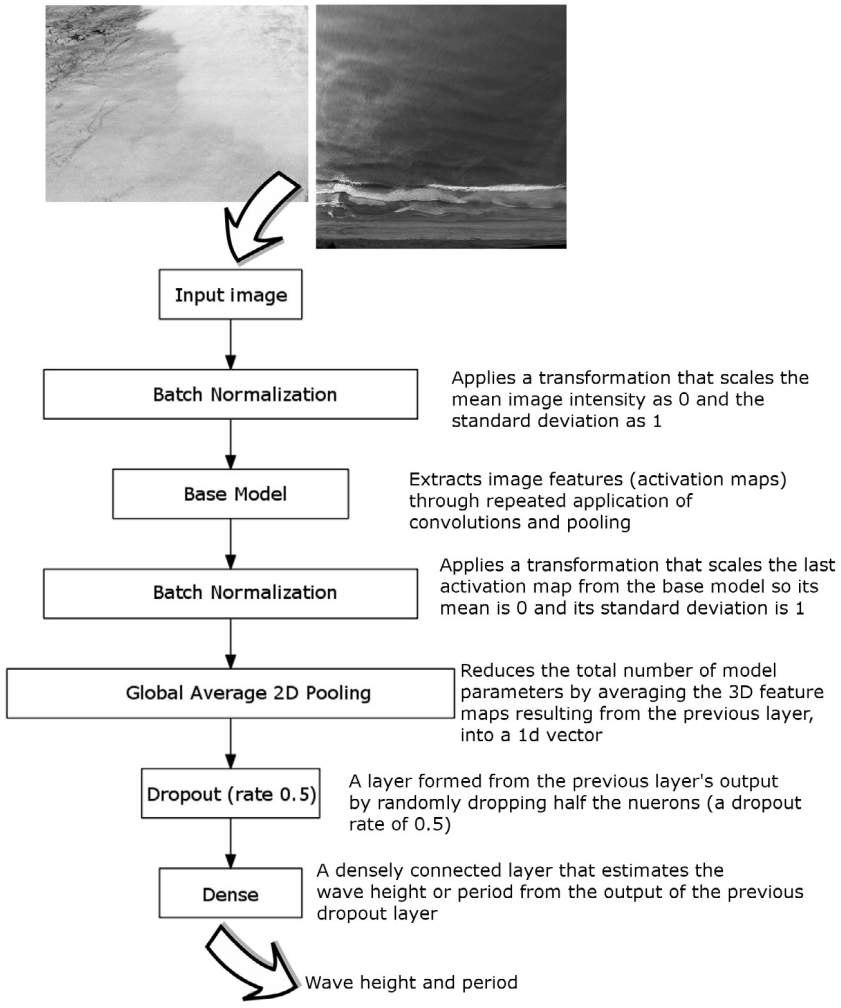

Fig. 7. Schematic of the generic technique used to estimate wave properties from input imagery.

Table 1
Details of the four model architectures used, specifying the number of parameters in each model component.

Layer	Model 1	Model 2	Model 3	Model 4
Batch normalization	4	4	4	4
Base model	MobilenetV1 (3,228,288)	MobilenetV2 (2,257,408)	InceptionV3 (21,802,208)	Inception-ResnetV2 (54,336,160)
Batch normalization	4096	5120	8192	6144
Global average pooling	0	0	0	0
Dropout	0	0	0	0
Dense	1025	1281	2049	1537
Total	3,233,413	2,263,813	21,812,453	54,343,845

predictions made by Inception-ResnetV2-based OWGs (with residual layers) tend to better those made by InceptionV3-based OWGs (without residual layers). Smaller batch sizes tend to result in more accurate OWGs.

An out-of-calibration validation, using images associated with wave heights or periods outside the range of values represented in the training data, showed that the ability for OWGs to predict the bottom 5% of low wave heights and the top 5% of high wave heights was reasonably good (Table 2). The mean across all four base models ranged between 22–29cm for the IR data, 18–30cm for the rectified RGB imagery, and 22–32cm for the oblique RGB imagery. The out-of-calibration errors were significantly larger for wave period (Table 3). The mean across all four base models ranged between 1.81–2.32 s for the IR data, 3.82–4.15 s for the rectified RGB imagery, and 1.98–2.07 s for the oblique RGB imagery.

4.1. Infrared imagery

Of the 16 OWGs trained to predict wave height (Fig. 8), the best overall performance was one based on MobileNetV1 with a batch size of 32 (Fig. 8B), with RMS error of 14 and 9 cm for within- and out-of-calibration-validation, respectively. Several other models based on MobileNetV1 or Inception-ResnetV2 had similar accuracy. RMS errors tend to increase with increasing batch size.

For wave period prediction (Fig. 9), there is less variability among base models and batch sizes. All wave period OWGs show a greater degree of scatter compared to OWGs for wave height. The best performing OWG overall was that based on Inception-ResnetV2 with a batch size of 128 (Fig. 9P). It produced an RMS error of 0.41 s within-calibration and 1.72 s out-of-calibration. However, the smallest model with the smallest batch size (Fig. 9A) performed almost as well within-calibration (0.43 s) and better out-of-calibration (1.65 s).

Table 2
Summary of model evaluation results for wave height. Root-mean-square error in wave height (m) for the three datasets, as a function of batch size and base model.

Oblique IR:					
	16	32	64	128	Mean:
MobilenetV1	0.15 (0.09)	0.14 (0.09)	0.16 (0.12)	0.31 (0.2)	0.19 (0.13)
MobilenetV2	0.73 (0.71)	0.57 (0.57)	0.52 (0.62)	0.56 (0.68)	0.6 (0.65)
InceptionV3	0.16 (0.12)	0.15 (0.08)	0.14 (0.09)	0.20 (0.14)	0.16 (0.11)
Inception-ResnetV2	0.14 (0.12)	0.17 (0.14)	0.15 (0.09)	0.20 (0.15)	0.17 (0.13)
Mean:	0.3 (0.27)	0.26 (0.22)	0.24 (0.23)	0.32 (0.29)	0.28 (0.25)
Orthomosaic RGB:					
	16	32	64	128	Mean:
MobilenetV1	0.1 (0.14)	0.12 (0.22)	0.16 (0.2)	0.21 (0.31)	0.15 (0.22)
MobilenetV2	0.12 (0.17)	0.26 (0.35)	0.32 (0.35)	0.32 (0.35)	0.26 (0.31)
InceptionV3	0.1 (0.21)	0.1 (0.15)	0.17 (0.21)	0.16 (0.29)	0.13 (0.21)
Inception-ResnetV2	0.09 (0.2)	0.08 (0.16)	0.1 (0.17)	0.16 (0.26)	0.11 (0.19)
Mean:	0.1 (0.18)	0.14 (0.22)	0.19 (0.23)	0.21 (0.3)	0.16 (0.23)
Oblique RGB:					
	16	32	64	128	Mean:
MobilenetV1	0.11 (0.23)	0.13 (0.21)	0.17 (0.24)	0.23 (0.35)	0.16 (0.26)
MobilenetV2	0.14 (0.21)	0.25 (0.24)	0.28 (0.34)	0.3 (0.4)	0.24 (0.3)
InceptionV3	0.16 (0.31)	0.12 (0.2)	0.2 (0.38)	0.16 (0.26)	0.16 (0.29)
Inception-ResnetV2	0.11 (0.23)	0.13 (0.23)	0.13 (0.18)	0.18 (0.27)	0.14 (0.23)
Mean:	0.13 (0.22)	0.16 (0.22)	0.2 (0.29)	0.22 (0.32)	0.18 (0.26)

Table 3
Summary of model evaluation results for wave period. Root-mean-square error in wave period (s) for the three datasets, as a function of batch size and base model.

Oblique IR:					
	16	32	64	128	Mean:
MobilenetV1	0.43 (1.65)	0.44 (2.13)	0.44 (1.78)	1.58 (3.52)	0.73 (2.27)
MobilenetV2	0.48 (1.9)	0.54 (2.07)	0.52 (2.33)	0.62 (1.89)	0.54 (2.05)
InceptionV3	0.47 (1.99)	0.42 (2.05)	0.42 (2.31)	0.41 (2.13)	0.43 (2.12)
Inception-ResnetV2	0.63 (1.69)	0.63 (2.1)	0.43 (2.06)	0.41 (1.72)	0.53 (1.89)
Mean:	0.5 (1.81)	0.51 (2.09)	0.45 (2.12)	0.76 (2.32)	0.56 (2.09)
Orthomosaic RGB:					
	16	32	64	128	Mean:
MobilenetV1	1.45 (3.91)	1.70 (3.83)	1.75 (3.94)	2.31 (4.3)	1.80 (3.99)
MobilenetV2	2.96 (4.57)	3.03 (3.92)	2.64 (4.25)	3.96 (5.0)	3.15 (4.43)
InceptionV3	1.47 (3.78)	0.84 (3.80)	1.12 (3.68)	1.25 (3.44)	1.17 (3.68)
Inception-ResnetV2	0.79 (3.91)	0.84 (3.72)	1.09 (3.79)	1.2 (3.88)	0.98 (3.83)
Mean:	1.67 (4.04)	1.6 (3.82)	1.65 (3.92)	2.18 (4.15)	1.78 (3.98)
Oblique RGB:					
	16	32	64	128	Mean:
MobilenetV1	1.41 (1.39)	2.17 (2.33)	2.07 (2.86)	1.99 (3.05)	1.91 (2.41)
MobilenetV2	2.67 (3.31)	2.48 (2.87)	3.65 (3.58)	3.36 (3.3)	3.04 (3.27)
InceptionV3	0.83 (1.76)	0.9 (1.71)	1.47 (2.1)	2.53 (3.36)	1.43 (2.23)
Inception-ResnetV2	0.86 (1.46)	0.81 (1.37)	1.03 (1.53)	1.34 (1.88)	1.01 (1.56)
Mean:	1.44 (1.98)	1.59 (2.07)	2.05 (2.52)	2.31 (2.9)	1.85 (2.37)

Overall, for both wave height and period, the prediction skill on 40% of the data suggests that the models do not overfit the training data, i.e. they generalize well to unseen data. There is little to separate the InceptionV3 and Inception-ResnetV2 base models. OWGs for wave height (Fig. 8) tend to perform significantly better out-of-calibration than OWGs for wave period (Fig. 9). Wave period models tend to over-predict the extremely low wave periods and under-predict the extremely high values. In general, there is no significant advantage in using either larger base models or larger batch sizes.

4.2. Visible-band imagery: rectified orthomosaics

Of the 16 OWGs trained to predict wave height (Fig. 10), the best overall performance was one based on MobileNetV1 with a batch size of 16 (Fig. 10A), with RMS error of 10 and 14 cm, respectively, for within- and out-of-calibration-validation. Like for the IR data, several other models (with the various batch sizes) based on Inception-ResnetV2 had similar accuracy, and RMS errors tend to increase with increasing batch size.

For wave period prediction (Fig. 11), there is more variability among base models and batch sizes compared to the IR dataset. However, like for the IR data, all wave period OWGs show a greater degree of scatter compared to OWGs for wave height. The best performing OWG overall was that based on Inception-ResnetV2 with a batch size of 16 (Fig. 9M), with RMS error of 0.79 s within-calibration, and 3.91 s out-of-calibration. The smallest models with the smallest batch sizes performed significantly worse within-calibration but very similar out-of-calibration. For this data, larger models clearly perform better for wave period.

Overall, for both wave height and period, the prediction skill on 40% of the data suggests that the models based on InceptionV3 and Inception-ResnetV2 base models do not overfit the training data, i.e. they generalize well to unseen data. OWGs for wave height (Fig. 10) tend to perform significantly better out-of-calibration than OWGs for wave period (Fig. 11). Models tend to over-predict the extremely low wave periods, and under-predict the extremely high values. InceptionV3 and Inception-ResnetV2 based models tend to either predict extremely high wave periods perfectly, or significantly under-predict them, with few values in between (Fig. 11I—P).

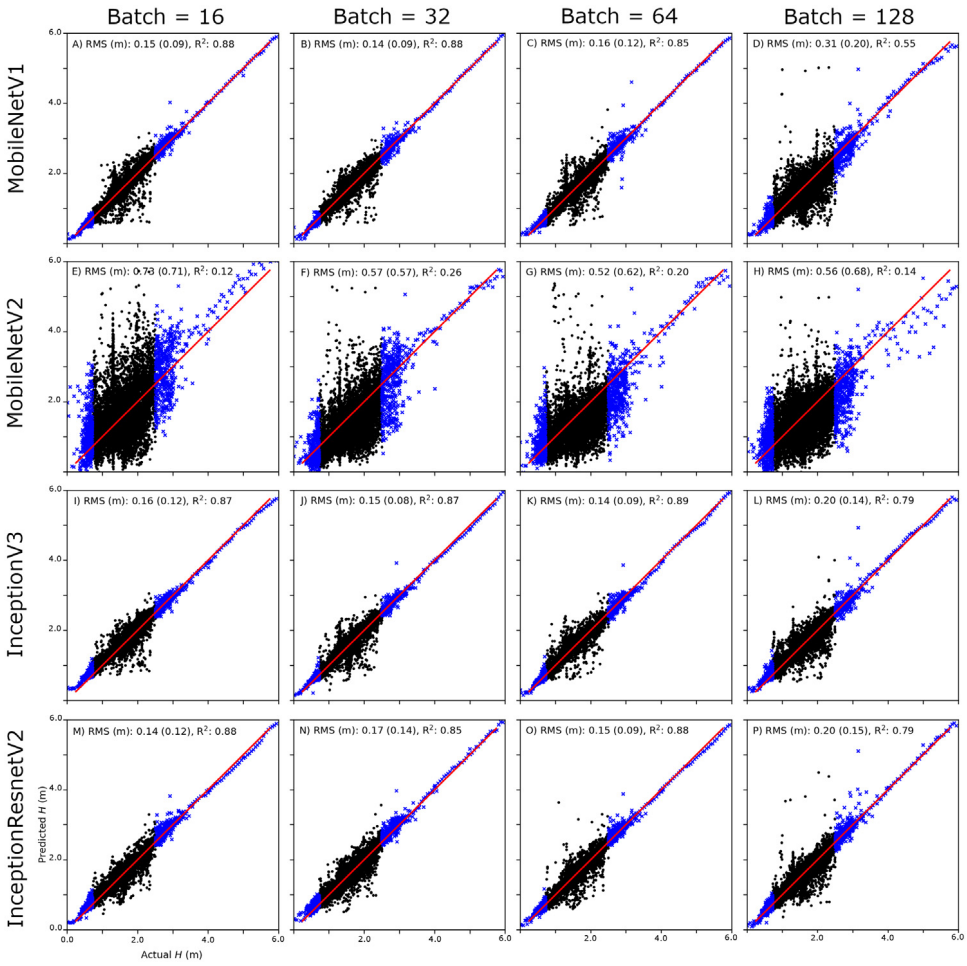

Fig. 8. Estimated versus observed wave height for the IR imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

4.3. Visible-band imagery: unrectified obliques

Of the 16 OWGs trained to predict wave height (Fig. 12), the best overall performance was achieved using MobileNetV1 with a batch size of 16 (Fig. 12A) and Inception-ResnetV2 with a batch size of 16 (Fig. 12M), and RMS error of 10 and 14 cm for within- and out-of-calibration-validation, respectively. Unlike for the IR and rectified RGB dataset, many more models (with the various batch sizes) based on InceptionV3 and Inception-ResnetV2 had similar accuracies. Like for the other datasets, RMS errors tend to increase with increasing batch size.

For wave period prediction from the orthomosaic RGB imagery (Fig. 13), like with the rectified RGB data, there is more variability among base models and batch sizes compared to the IR dataset. However, like for all datasets considered here, all wave period OWGs show a greater degree of scatter compared to wave height OWGs. Like for the rectified RGB imagery, the best performing OWG overall was that based on Inception-ResnetV2 with a batch size of 16 (Fig. 13M). It yielded RMS errors of 0.86 s within-calibration and 1.46 s out-of-calibration. The out-of-calibration scores in general are significantly better for the oblique imagery (Fig. 13) compared with the rectified imagery (Fig. 11). The smallest models with the smallest batch sizes performed significantly worse than the larger models with larger batch size within-calibration, but similarly out-of-calibration. For this oblique RGB dataset, the larger models and smallest batch sizes most clearly perform better for wave period.

Overall, for both wave height and period, the prediction skill on 40% of the data suggests that the models based on Inception-ResnetV2 base models do not overfit the training data, i.e. they generalize well to unseen data. Like for the other datasets, OWGs for wave height (Fig. 12) tend to perform significantly better out-of-calibration than OWGs for wave period (Fig. 13). However, unlike the models for rectified imagery, models for oblique imagery do not tend to suffer from over-predicting the extremely low wave periods and under-predicting the extremely high values.

5. Discussion

5.1. OWG performance

Based on tests on independent data (40% of all data), all OWGs performed reasonably well at estimating both wave period and height, generalizing well beyond the training set (60% of all data). Within-calibration OWG accuracies were sensitive to both the choice of model architecture and batch size. Generally speaking, where within-calibration errors were relatively low, so too were out-of-calibration accuracies. The OWG technique as presented here is suitable for predicting quantities outside of the range of values represented within the training data, however that ability was generally much better for wave height than for wave period. In general, therefore, we recommend that training datasets should be large enough that include example images from extreme events, especially for extremely long or short wave periods. For longer-term OWG deployment, larger data sets may

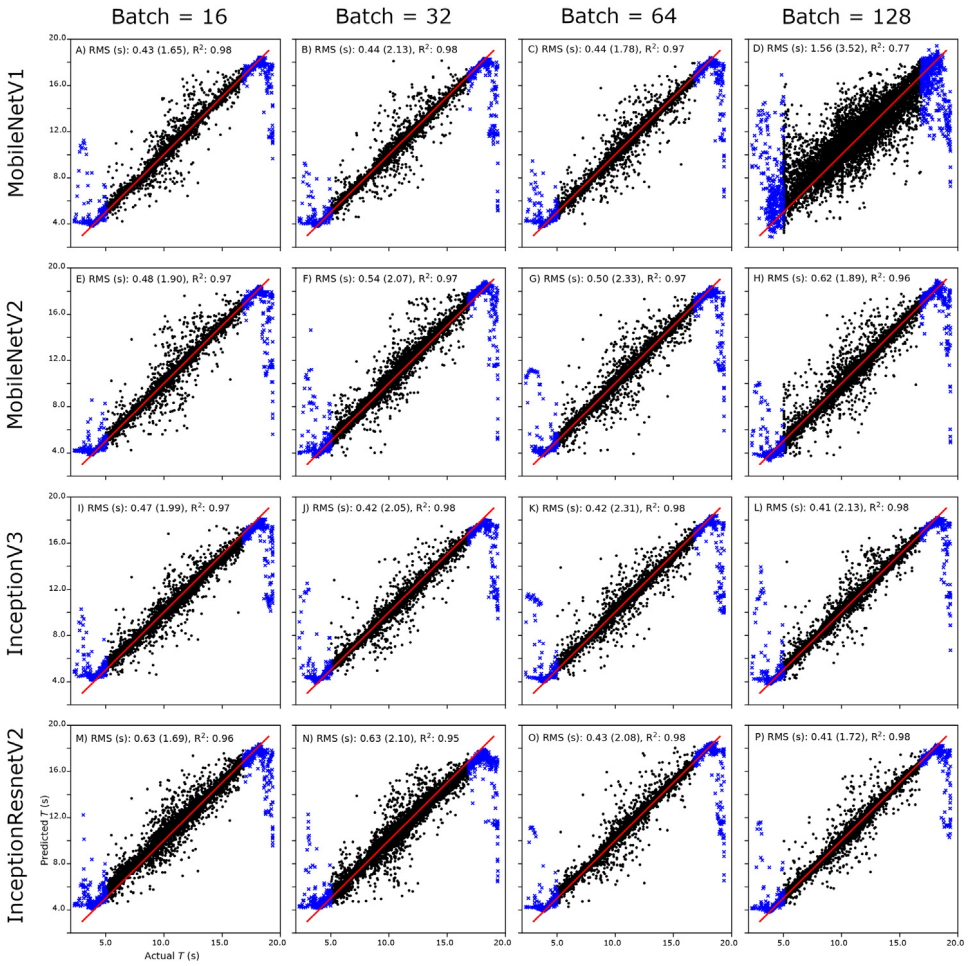

Fig. 9. Estimated versus observed wave period for the IR imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

allow for less well-balanced test/training splits, perhaps as high as 75/25% test/train split or even higher. That OWG performance does not generally improve with batch size suggests that performance should not scale with computer resources. Large datasets should be trained using relatively small base models such as MobilenetV1. Limited trials with larger input image sizes suggest that OWG performance would increase with larger images, given sufficient GPU memory, but at the cost of significantly longer model training times.

By comparing the outputs of optical wave gauges built around base models MobilenetV1 and MobilenetV2, we can examine the effectiveness of residual layers in a relative small CNN model, and by comparing gauges built around InceptionV3 and Inception-ResnetV2, the effectiveness of residual layers in a relatively large CNN model can be assessed. For all three sets of imagery, it can be concluded that the presence of residual layers in a basic Mobilenet architecture makes OWGs generally less accurate, but the opposite is true for the basic Inception architecture. In general, based on overall estimation for all three datasets and both variables, the models based on the smallest and simplest model, MobilenetV1, were optimal for wave height. Those models based on Inception-ResnetV2 were generally optimal for wave period. This suggests that there could be an even smaller (computationally more efficient) as-yet undiscovered optimal feature extractor for the present task. Generally, the number of neurons in the final dense layer of the OWG, hence the resolving power of the network, does not seem to be an important factor; InceptionV3 has the largest number of these neurons (Table 1) but does not tend to result in the greatest accuracy. Overall, the optical wave gauge built upon the MobilenetV1 base model with 3.2M parameters, might be a more efficient and parsimonious predictor than Inception-ResnetV2 with 54.3M parameters. MobilenetV1, as the name suggests, was designed to be deployed on mobile devices; small numbers of parameters equate to relatively small files that contain model checkpoints (a consideration for mobile device applications) and more efficient use of GPU memory, which means faster training times and larger batch sizes to be held in (relatively expensive) GPU memory. MobilenetV1 could therefore be a good starting point for fine-tuning architectures in order to find the most parsimonious and/or generally applicable model base architecture.

The OWG estimates the characteristics of the wave field integrated throughout the image rather than individual waves within the image. Both in infrared images of breaking waves (Fig. 3) and conventional photographic images of entire fields of propagating and breaking waves (Figs. 5 and 6), similar patterns repeat in different parts of an image, so they exhibit a high degree of spatial stationarity. We suggest that our approach worked well because CNNs capture and exploit image stationarity (Krizhevsky et al., 2012). This means that features that are useful in one region are also likely to be useful for other regions. In practice, having learned relevant features over small patches sampled randomly from the larger image, the CNN then applies this learned small feature detector everywhere in the image. These features are then preserved by using maxima when pooling. After filtering input imagery using convolutions, maximum pooling of those convolved features activates the same features even while the image undergoes progressive downsizing.

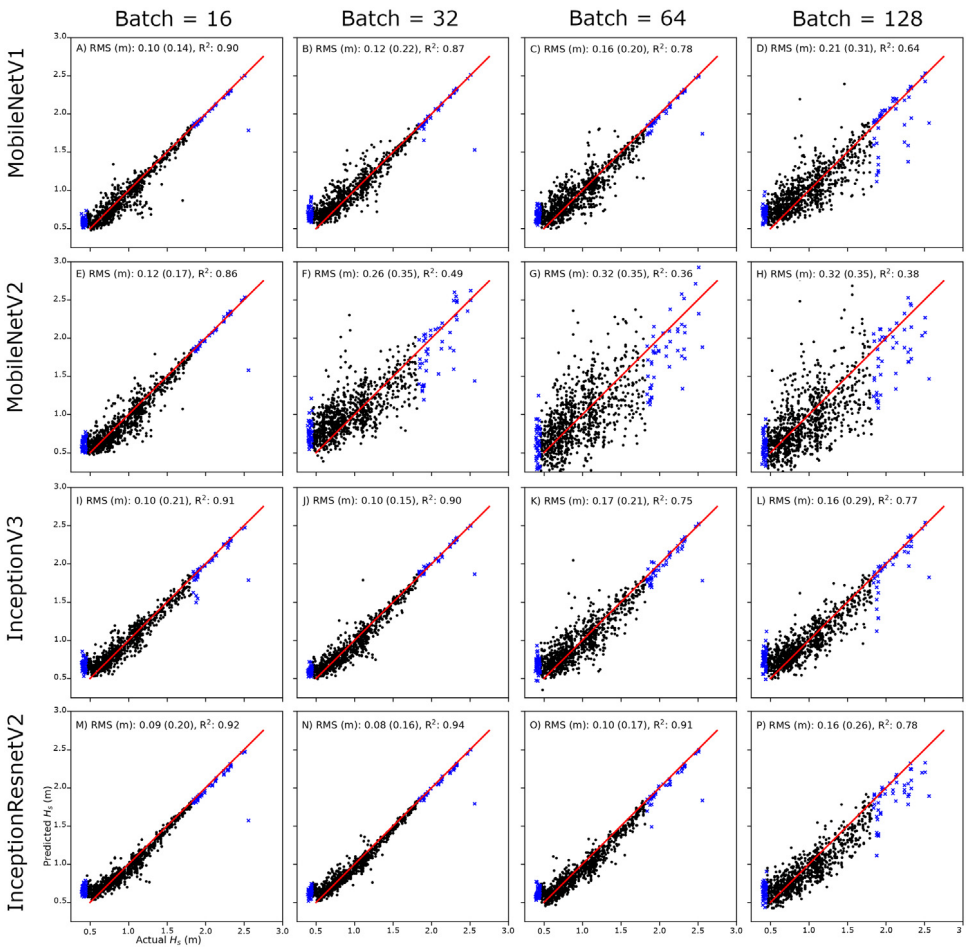

Fig. 10. Estimated versus observed H_s for the rectified visible band nearshore imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

By comparing errors across two oblique RGB sets of imagery with an irregular spatial footprint, and one rectified set of images with a regular spatial footprint, we conclude that image rectification, or otherwise any knowledge of camera geometry, is not required for successful application of the technique. In this respect the comparison between the rectified and oblique RGB Argus imagery is most illuminating, since they are trained using the same wave data. Based on summary mean RMS errors reported in Tables 2 and 3, we conclude that there is only a marginal advantage to rectifying imagery for the purposes of wave height; Inception-ResnetV2 models trained with a batch size of 16 are accurate for within-calibration data to within 9 and 11 cm for rectified and oblique RGB imagery, respectively (Table 2). Out-of-calibration errors are 20 and 23 cm, respectively. However, for wave period we conclude that not rectifying imagery offers a significant advantage to out-of-calibration wave period estimation.

5.2. Image feature extraction

It is instructive to visualize the features extracted by the network. To do so, we display the mean output (so called 'activation') over the last convolutional block in the Inception-ResnetV2 feature extractor, using the rectified visible band imagery. Fig. 14 shows a selection of images associated with increasing H_s, randomly selected from eight wave height bins in the record. Fig. 15 shows a selection of images associated with eight increasing T_p bins. Column A in Figs. 14 and 15 shows the grayscale image inputs; column B shows the corresponding average feature maps extracted using weights learned using the Imagenet (Deng et al., 2009) dataset that is commonly used for transfer learning (Buscombe and Carini, 2019), and column C shows the average feature maps extracted using weights learned on our data.

Relatively bright pixels in Figs. 14 and 15 indicate areas that the network has decided are relatively important for estimating wave height or period. Note that the two dimensional activations shown in Figs. 14 and 15 are mean values over the last three-dimensional set of activations. The lack of consistent or physically interpretable spatial pattern in features extracted using models with Imagenet weights clearly demonstrate that 'transfer learning', where a model trained on one task is re-purposed on a second related task (Buscombe and Ritchie, 2018; Buscombe and Carini, 2019), would not be effective for optical wave gauging. In other words, OWGs are only successful because they are trained end-to-end, using a cost function to tune the weights of the network to optimize feature extraction for a specific quantity with a specific camera field-of-view.

Although the validation data is measured offshore (towards the seawards extent of the image), the most diagnostic image features for estimating H_s are nearshore; the OWG optimized for wave height clearly uses the surf (and, to a lesser extent, swash) zones (Fig. 14) to make predictions, and that is some basic function of number, location, alongshore extent and, perhaps most importantly, the width of bright pixels near the shoreline (i.e. surf zone width). The width of the surf zone as the most diagnostic feature makes physical sense for a saturated inner surf zone on dissipative beaches, where the incident wave height is controlled by the local water depth (Thornton and Guza, 1982). Therefore the OWG could be sensitive to the beach's morphodynamic state, tide, and relative shoreline position. For an operational setting,

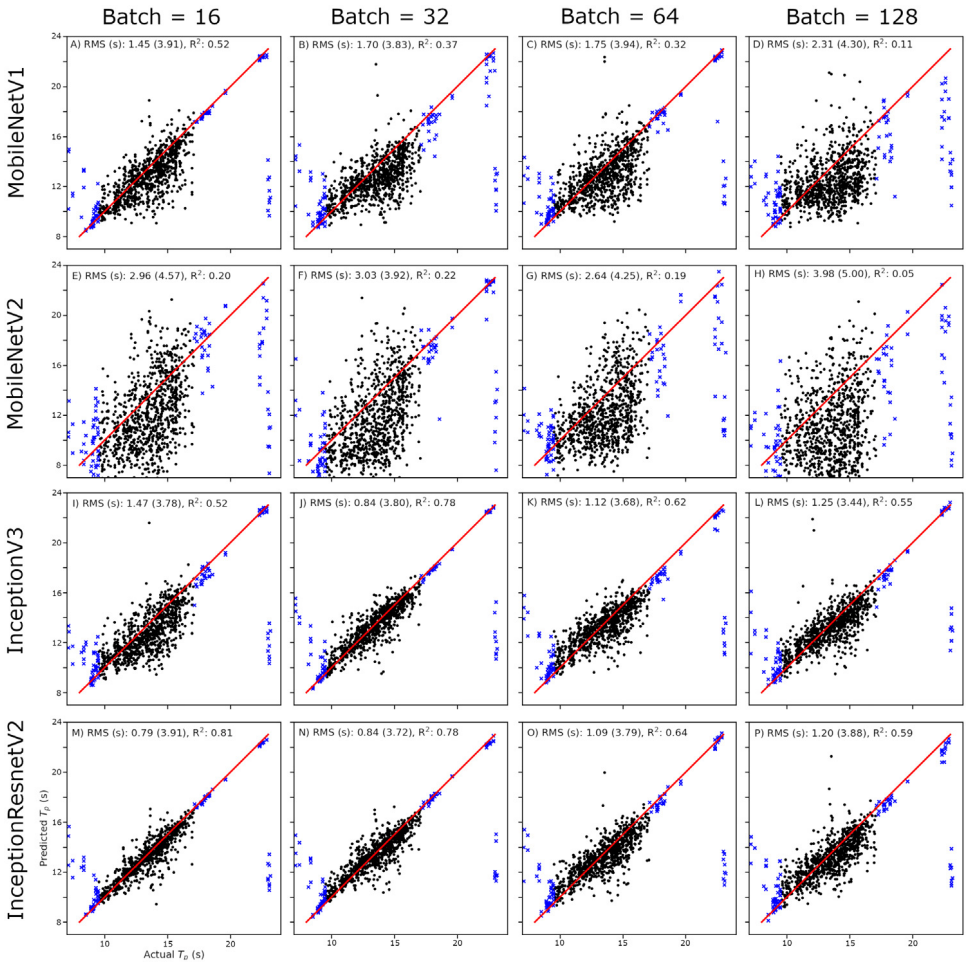

Fig. 11. Estimated versus observed T_p for the rectified visible band nearshore imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

OWGs should be trained using data from multiple seasons, perhaps multiple years, otherwise there would be systematic model deviation from measurements when the beach underwent a change in morphodynamic state. Some bright features in offshore areas of the average feature maps suggest that the OWG might, perhaps secondarily, be using the disparity of shoaling wave face elevation as they are distorted onto the horizontal rectified plane. Physically, the base model could be receptive to an initial wave height decrease as it shoals, and its subsequent increase shortly before breaking, but that is just speculation.

The same analysis was performed on the model optimized for wave period (Fig. 15). In this case, many more features from the images are preserved in offshore locations, indicating much more of the image is important for the T_p prediction than for the H_s prediction. In contrast to the features extracted for estimating H_s (Fig. 14), surfzone areas are not important for T_p prediction. Instead, the OWG is indicating sensitivity to areas outside of the surfzone, where unbroken waves are visible. It is reasonable that the wave period information is derived from the wave length observed in this region, which is a linear function of wave period in shallow water. The model might also be extracting features in this region associated with the wavelength decreasing up to breaking as waves shoal (Sakai and Battjes, 1980).

5.3. Possible future directions

It is remarkable that the same framework, not optimized for any particular dataset, can predict both wave height and period with high accuracy, despite the differences in electromagnetic band, perspective, and scale. This suggests it might prove transferable between sites, scales and camera platforms; a claim that should be investigated further. That said, models do not predict other models training data well, which suggests that while the model framework might be universal, it is not picking up anything universal in the data, and the network weights for the three sets of models are correspondingly very different. The success of the technique is inherently tied to the invariance of the stationary camera's pose and viewpoint.

'Out-of-calibration' tests indicate that the model framework is capable of modeling the input data in a sufficiently general way to estimate well beyond the range of values represented in the training data. However, the performance was sensitive to base model, and was typically worse for wave period. This may suggest that extracted features are highly specific to narrow ranges of wave periods, or it might be the result of a class imbalance problem —there are far fewer examples of extremely small and extremely large wave periods than mid-size periods. Our use of stratified random sampling to draw batches of training and testing images from ten monotonically increasing wave period bins, designed to avoid introducing any frequency bias associated with selecting wave period magnitudes based on their relative proportion within the training image set, was perhaps misguided. For example, perhaps we should have sampled exponentially rather than linearly increasing bins. Beyond simply acquiring more training data to increase the likelihood of sampling extreme events, there are other strategies that could be explored in future work, such as ways to balance the loss function (Johnson and Khoshgoftaar, 2019) based on the relative frequency of wave heights and/or periods. This would

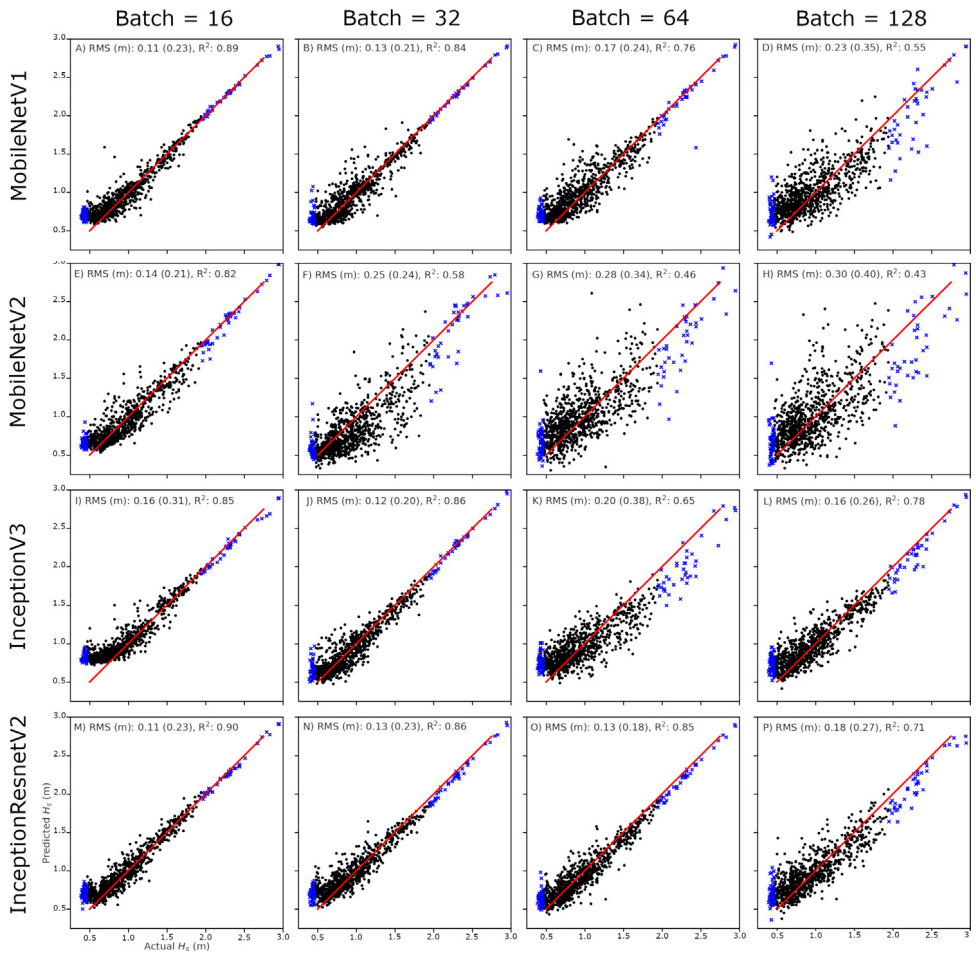

Fig. 12. Estimated versus observed H_s for the oblique visible band nearshore imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

impose an additional cost on the model for making mistakes on the minority values during training, deliberately biasing the model to pay more attention to the extreme values. The effects of dropout and data augmentation (i.e. too much regularization) on the out-of-calibration predictive skill of the models should also be explored in future work.

The inability of OWG models to predict wave periods greater than those represented in the IR training set might be related to the physical size of the image, which was only approx 20 × 20 m. Linear wave theory would predict an approximately 50-m wavelength for a 15-second period wave, therefore the sub-wavelength field-of-view possibly did not see the range of image textures associated with wave geometry and sea-surface temperature patchiness diagnostic of those larger period waves. Future work should therefore examine scale-dependency in OWG estimates.

There was generally a significant advantage in the use of MobileNetV1 and Inception-ResnetV2 as base models. This motivates focusing efforts on the discovery of the optimal generic image feature extractors from those previously proposed in the literature for generic image classification, such as VGG (Simonyan and Zisserman, 2014) or Xception (Chollet, 2017). Alternatively, the general skill of the smallest model, MobileNetV1, may suggest that research into smaller, less complex feature extractors is warranted in order to discover the smallest network that still provides reasonable wave height or period estimates. The value of a given proposed feature extractor model should be evaluated on its ability to predict wave statistics of magnitude larger or smaller than those represented within the training data (out-of-calibration validation). Further, the OWGs implemented here were not optimized for any dataset, and could be further optimized for individual datasets. This optimization could involve a more exhaustive exploration of different base models, network layers, and hyperparameters, which should be the subject of future work.

The OWG models presented here are for site-specific monitoring; therefore they are designed to be used to estimate wave quantities by taking as input imagery that is the same field-of-view as the training dataset. That implies they are tied to a particular geographic location. Further, the OWG technique estimates a scene-averaged wave statistic, and is therefore not intended to measure each wave within a scene, for example to obtain cross-shore profiles of wave quantities. However, further work may explore this possibility using a cross-shore instrumented array to train models to estimate wave quantities in specific portions of imagery. In this study, the stationary cameras experienced very slight movement; with variation in the oblique field of view less than 2 pixels for a target at 600 m distance from the camera, which translates to less than 5 m in the along-camera axis in the rectified domain. Slight movement such as this is barely detectable in the oblique field of view and does not impact model results. However, if the cameras did move significantly (enough to change the entire vantage) that would affect model results. Therefore, the degree to which models are sensitive to camera movement is unknown and should be studied further. Future work should also explore alternative model architectures that extract relevant image features from generic imagery of the surf zone. This should possibly be tackled in stages, by first optimizing OWGs to predict at one site reliably regardless of morphological change; then

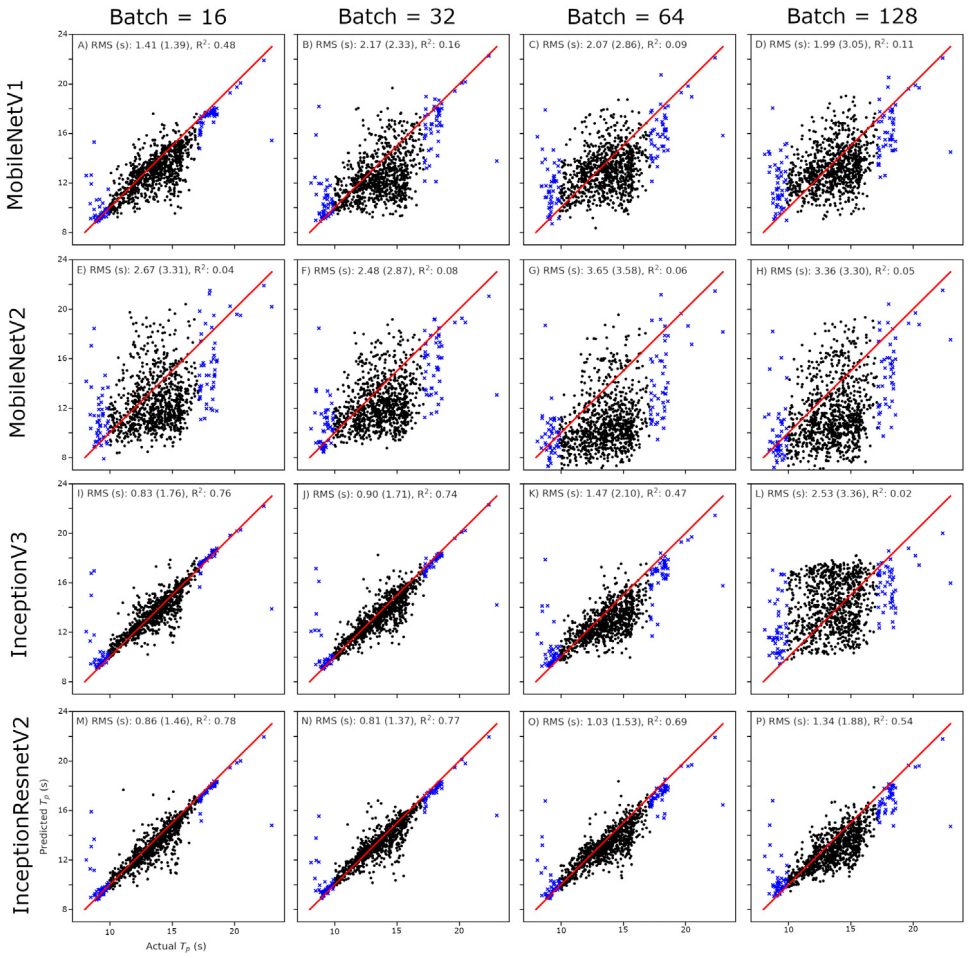

Fig. 13. Estimated versus observed T_p for the oblique visible band nearshore imagery test dataset. From left to right, batch size increases from 16 to 32 to 64 and finally 128. From top to bottom, OWGs based on MobilenetV1; MobilenetV2; InceptionV3; and Inception-ResnetV2. Lines show 1:1 correspondence. RMS error refers to the within-calibration test samples (black dots). The RMS error in parentheses refers to the out-of-calibration samples (blue crosses).

overcoming any sensitivities to image scale, vantage, and perspective; and finally exploring sensitivities to greater variation in beach morphologies, grain sizes, and surf zone characteristics.

The OWG can be trained for both individual wave height and period (such as the IR imagery), and statistical quantities like significant wave height and peak wave period (such as the nearshore Argus imagery). An advantage of the OWG technique is that the image does not need to be rectified onto a regular grid, therefore ground control points do not need to be obtained and image geometries do not need to be computed and applied. Another advantage is that it estimates a wave statistic from a single image. Therefore, assuming errors are random and can be reduced through averaging, one strategy for reducing error of each estimate is to over-sample in space (i.e. increase the number of cameras and models) or time (i.e. increase the sample frequency of the camera) then average over a short series of those high-frequency estimates for a more accurate, albeit lower-frequency, estimate. For example, such averaging in time might better safeguard against an OWG under-predicting wave height in apparent lulls between sets. Infrared imaging might be useful for optical wave gauging at night. In addition to providing a low-cost routine for monitoring waves (Fig. 16), techniques such as this could help validate larger-scale buoy-driven numerical wave models (O'Reilly et al., 2016; Crosby et al., 2017), HF radar inversions (Gurgel et al., 1999), and X-band radar observations (Dankert and Rosenthal, 2004) in numerous nearshore locations. Wave height or period estimation time is primarily a function of model size (number of parameters), increasing from 21 ms per image using MobilenetV1 as the base model, to 63 ms using Inception-ResnetV2, on a modest 2.2 GHz CPU. These sub-decisecond model execution times suggest 'real-time' wave estimation would be possible.

For an operational setting, OWGs should be trained using data from multiple seasons, perhaps multiple years, to capture a large range of magnitudes including extreme events, and to capture covariance between wave quantities and any surf zone morphologies such as seasonal changes in sandbar systems, 3D morphology from the presence of rip channels, and meso- to megacuspate features. Given that the features used to estimate wave height and period differ significantly, we suggest that OWGs trained separately for individual quantities, such as here, are likely to be more accurate than OWGs trained to predict multiple quantities simultaneously using a common set of extracted features. However, this should be further explored, by comparing against models trained to predict multiple quantities from a single image. A better test of the technique than that presented here would be a training period consisting of several months to years, followed by a subsequent period of equal or longer duration, which would test how well the model captures variability over multiple timescales, including co-variability in beach morphology and waves, as represented in model-extracted image features. In future developments, serial correlation in the data itself might also be exploited to greater effect.

Almost all applications of DCNNs for data-driven prediction to date have been made with images obtained using incoherent natural light or radiation. Within coastal oceanography, measurements made with coherent images are also common, such as those obtained by holography (Davies et al., 2015), radar (Holman and Haller, 2013), or ultrasound (Thorne and Hurther, 2014), and based on the results

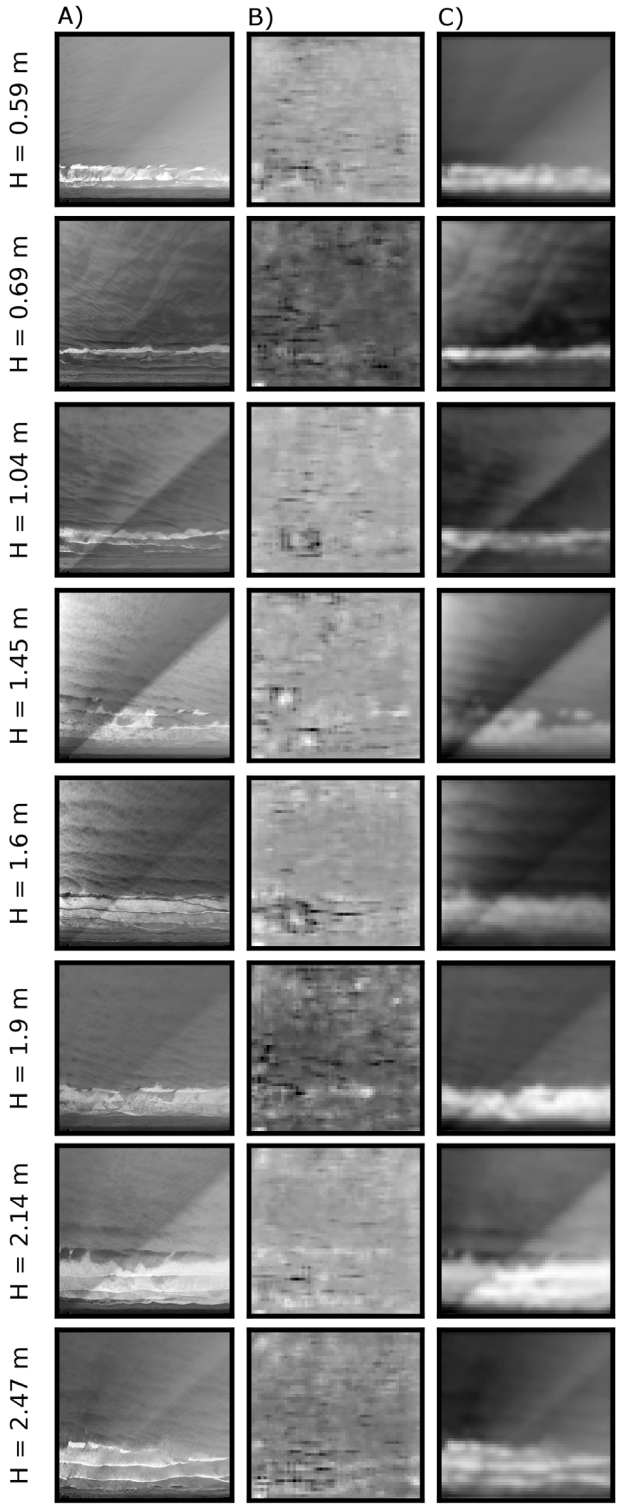

Fig. 14. From left to right: nearshore images associated with some observed H_s (A), the average of the corresponding features extracted by the Inception-ResnetV2 weighted using weights learned from the Imagenet dataset (B) and weighted using the weights learned in this study (C). Relatively bright pixels indicate areas that the network has decided is relatively important to estimate wave height.

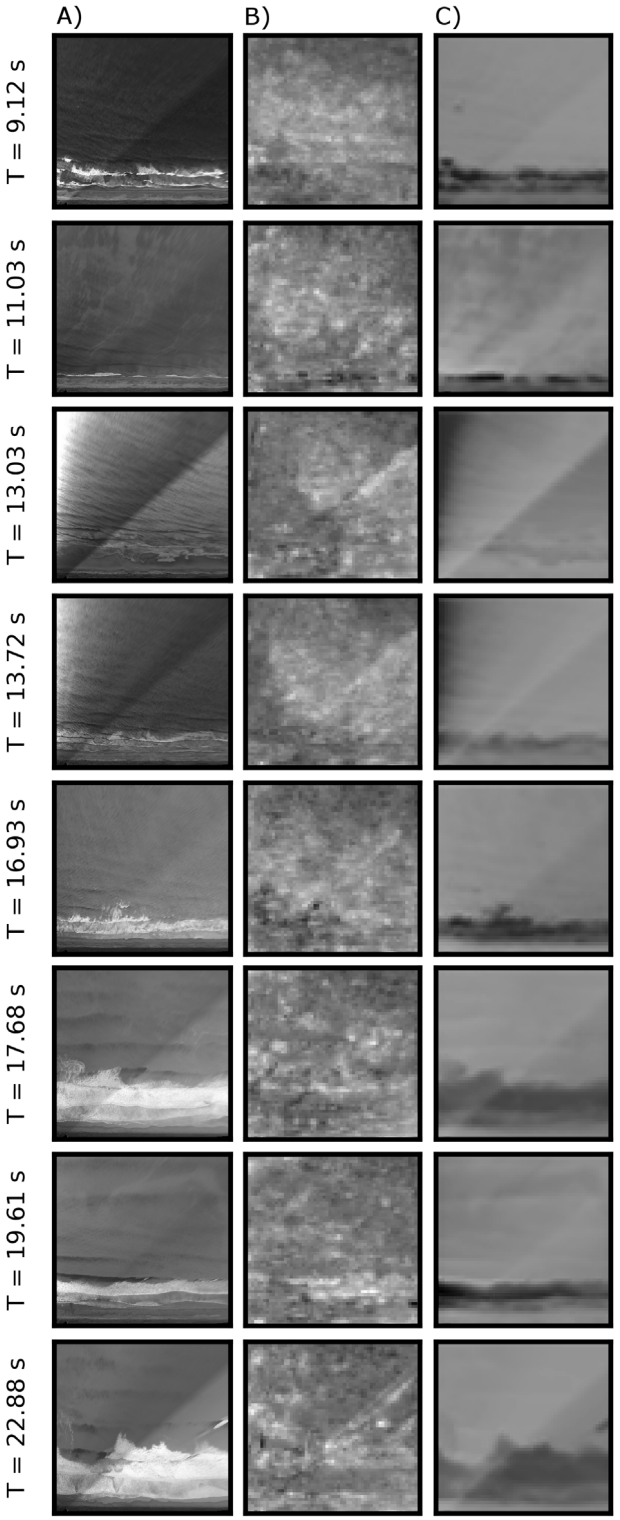

Fig. 15. From left to right: nearshore images associated with some observed T_p (A), the average of the corresponding features extracted by the Inception-ResnetV2 weighted using weights learned from the Imagenet dataset (B) and weighted using the weights learned in this study (C). Relatively bright pixels indicate areas that the network has decided is relatively important to estimate wave period.

presented here, might also be amenable to data-driven estimation of physical quantities using deep convolutional neural networks. The generally high signal-to-noise ratios in such imagery might further suggest the utility of such an approach.

6. Conclusions

This proof-of-concept study has demonstrated that, given sufficient images with paired wave data, it is possible to estimate wave height

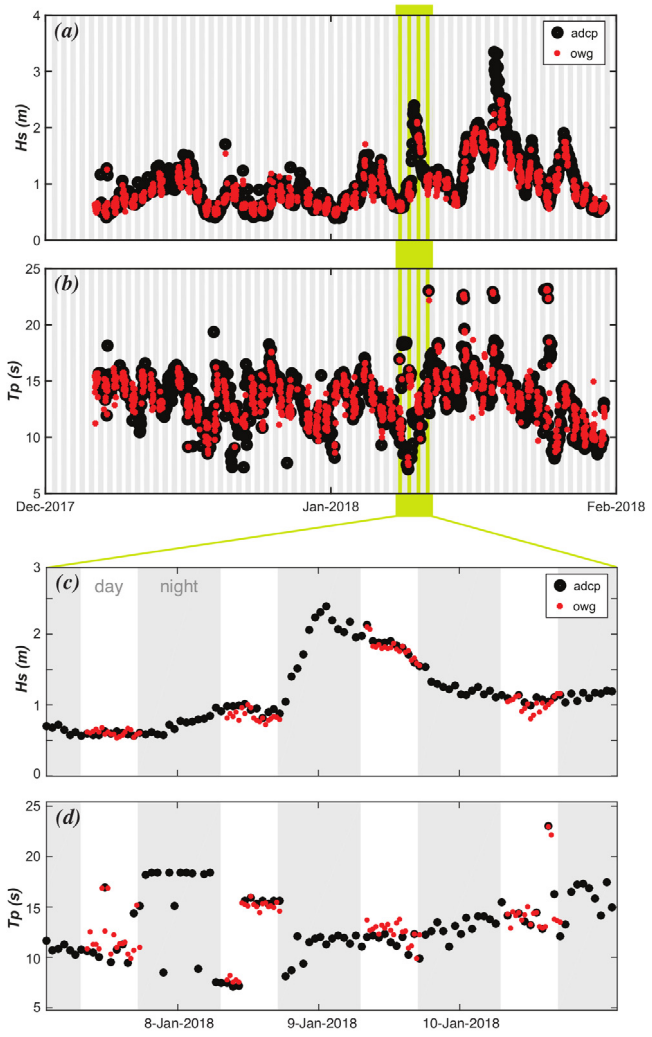

Fig. 16. Time-series of significant wave height, H_s (A) and peak wave period, T_p (B), as measured using ADCP (black markers) and OWG (red markers) at Sunset State Beach. Panels (C) and (D) shows a few days before, during, and after a moderately sized wave event. The alternating light/dark panel shading indicates day and night.

or period from a single image of waves, using a deep neural network model framework trained to a specific site and viewpoint/field-of-view. The model framework, called an Optical Wave Gauge or OWG can be trained for both individual wave height and period, and statistical quantities like significant wave height and peak wave period. We have demonstrated this concept using rectified and oblique RGB visible-band imagery, and oblique infrared (IR) imagery. Therefore our results strongly suggest that knowledge of specific camera geometries is not required for successful application of the method.

The best performing OWG on the IR dataset achieved RMS errors of 0.14 (0.08) m and 0.41 (1.65) s (values in parentheses are for out-of-calibration samples), for height and period respectively, capturing up to 98% of the variance in these quantities. The best performing OWG on the visible band rectified dataset achieved RMS errors of 0.08 (0.14) m and 0.79 (3.44) s for height and period, respectively. The same values for the oblique RGB imagery were 0.11 (0.18) m and 0.81 (1.37) s for height and period, respectively. The prediction skill on 40% of the data suggests that the models do not overfit the training data, i.e. they generalize well to unseen data. Both wave height and period estimates are somewhat sensitive to choice of base model; OWGs based on either MobileNetV1 or Inception-ResnetV2 tend to perform best. Smaller batch sizes tend to result in more accurate OWGs. However, operational deployments of this model framework might prove that the size and quality of available training data may be more important than specific feature extractor model or model training hyperparameters such as batch size.

OWGs for wave height tend to perform slightly better (both within- and out-of-calibration) than OWGs for wave period. Generally, most models perform reasonably well at predicting outside of the range of values represented in the training data, especially for wave height. However, OWGs tend to over-predict the extremely low wave periods and under-predict extremely high periods. Application of this method should therefore use a training dataset that captures the full variability of desired wave parameters, or at least should not be used to predict outside the magnitudes represented in the training data. Ways in which model training could be modified to result in OWGs that better estimate wave properties from images associated with extreme values outside of the values represented in the data used to train the OWG, or 'out-of-calibration' estimation, were discussed.

Applications of deep learning to geophysical imagery have so far only been made in a handful of experimental contexts. We believe this study to be the first application of deep learning to ocean wave characterization. It is also one of the first applications of a machine learning algorithm that can learn useful representations of features directly from monochrome geophysical imagery without feature extraction or selection. As such, this study opens several research avenues for the further investigation of models based on deep learning for the reconstruction of geophysical dynamics from remotely sensed data acquired from ground-based instruments.

Once optimized, this technique might compliment existing remote sensing techniques for nearshore wave monitoring. We hope and expect that new technologies, such as presented here, will inspire the future development of technically and operationally feasible data-driven observations of nearshore hydrodynamics. Future work will include: estimating model skill for wave conditions beyond the conditions within the training data; training and evaluating models trained over significantly different bathymetric beach states (i.e. winter vs summer profile, barred beaches, cusps, etc.); adaptation for reflective beaches and for macrotidal environments; exploring the possibility of models that can be transferred to other locations; and exploring the possibility of real-time gauging onboard an Argus station.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

RJC and CCC thank 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. RJC and CCC were funded by National Science Foundation, USA grant number 1736389 and ONR, USA grant number N000141010932. SRH and JAW thank Kurt Rosenberger, Jenny White, Tim Elfers, Gerry Hatcher and the USGS PCMSC MARFAC staff for field support; Coastal Imaging Research Network (CIRN) for software support; Funded by USGS, USA Remote Sensing Coastal Change project. DB received funding from Earth Science Information Partners (ESIP). Thanks to Chris Sherwood, Zafer Defne, Andrew Stevens and two anonymous reviewers for their suggestions. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. This study is fully reproducible using data and code available at https://github.com/dbuscombe-usgs/OpticalWaveGauging_DNN.

References

Aarninkhof, S.G., Ruessink, B.G., 2004. Video observations and model predictions of depth-induced wave dissipation. IEEE Trans. Geosci. Remote Sens. 42 (11), 2612–2622.

Allard, R., Dykes, J., Hsu, Y., Kaihatu, J., Conley, D., 2008. A real-time nearshore wave and current prediction system. J. Mar. Syst. 69 (1–2), 37–58.

Almar, R., Larnier, S., Castelle, B., Scott, T., Floc'h, F., 2016. On the use of the radon transform to estimate longshore currents from video imagery. Coastal Eng. 114, 301–308.

Baldock, T.E., Moura, T., Power, H.E., 2017. Video-based remote sensing of surf zone conditions. IEEE Potentials 36 (2), 35–41.

Benetazzo, A., 2006. Measurements of short water waves using stereo matched image sequences. Coastal Eng. 53 (12), 1013–1032.

Buscombe, D., 2019. Sedinet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry. Earth Surf. Process. Landforms http://dx.doi.org/10.1002/esp.4760.

Buscombe, D., Carini, R.J., 2019. A data-driven approach to classifying wave breaking in infrared imagery. Remote Sens. 11 (7), 859.

Buscombe, D., Ritchie, A., 2018. Landscape classification with deep neural networks. Geosciences 8 (7), 244.

Carini, R.J., Chickadel, C.C., Jessup, A.T., Thomson, J., 2015. Estimating wave energy dissipation in the surf zone using thermal infrared imagery. J. Geophys. Res. Oceans 120 (6), 3937–3957.

Chickadel, C., Holman, R.A., Freilich, M.H., 2003. An optical technique for the measurement of longshore currents. J. Geophys. Res. Oceans 108 (C11).

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proc. CVPR IEEE, pp. 1251–1258.

Chollet, F., et al., 2015. Keras. https://keras.io.

Crosby, S.C., Cornuelle, B.D., O'Reilly, W.C., Guza, R.T., 2017. Assimilating global wave model predictions and deep-water wave observations in nearshore swell predictions. J. Atmos. Ocean. Technol. 34 (8), 1823–1836.

Dankert, H., Rosenthal, W., 2004. Ocean surface determination from X-band radar-image sequences. J. Geophys. Res. Oceans 109 (C4).

Davies, E.J., Buscombe, D., Graham, G.W., Nimmo-Smith, W.A.M., 2015. Evaluating unsupervised methods to size and classify suspended particles using digital in-line holography. J. Atmos. Ocean. Technol. 32 (6), 1241–1256.

De Vries, S., Hill, D., De Schipper, M., Stive, M., 2011. Remote sensing of surf zone waves using stereo imaging. Coastal Eng. 58 (3), 239–250.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: Proc. CVPR IEEE. IEEE, pp. 248–255.

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: 9th IAPR Workshop PPRS. IEEE, pp. 1–6.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning, vol. 1. MIT Press, Cambridge.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2017. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377.

Gurgel, K.-W., Antonischki, G., Essen, H.-H., Schlick, T., 1999. Wellen radar (WERA): a new ground-wave HF radar for ocean remote sensing. Coastal Eng. 37 (3–4), 219–234.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. CVPR IEEE, pp. 770–778.

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 J. Oceanic Eng. 22 (1), 81–92.

Holman, R.A., Guza, R., 1984. Measuring run-up on a natural beach. Coastal Eng. 8 (2), 129–140.

Holman, R., Haller, M.C., 2013. Remote sensing of the nearshore. Annu. Rev. Mar. Sci. 5, 95–113.

Holman, R.A., Sallenger, A.H., Lippmann, T.C., Haines, J.W., 1993. The application of video image processing to the study of nearshore processes. Oceanography 6 (3), 78–85.

Holman, R.A., Stanley, J., 2007. The history and technical capabilities of argus. Coastal Eng. 54 (6–7), 477–491.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. ArXiv preprint arXiv:1704.04861.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv preprint arXiv:1502.03167.

Irish, J.L., Wozencraft, J.M., Cunningham, A.G., Giroud, C., 2006. Nonintrusive measurement of ocean waves: Lidar wave gauge. J. Atmos. Ocean. Technol. 23 (11), 1559–1572.

Jessup, A., Zappa, C., Loewen, M., Hesany, V., 1997a. Infrared remote sensing of breaking waves. Nature 385 (6611), 52.

Jessup, A., Zappa, C.J., Yeh, H., 1997b. Defining and quantifying microscale wave breaking with infrared imagery. J. Geophys. Res. Oceans 102 (C10), 23145–23153.

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. Geophys. Res. Lett. 45 (8), 3706–3716.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6 (1), 27.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. ArXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Adv. Neur. In. pp. 1097–1105.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436.

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 Geosci. Remote Sens. 14 (3), 354–358. http://dx.doi.org/10.1109/LGRS.2016.2643000.

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. Limnol. Oceanogr-Meth. 16 (12), 814–827.

O'Reilly, W., Olfe, C.B., Thomas, J., Seymour, R., Guza, R., 2016. The California coastal wave monitoring and prediction system. Coastal Eng. 116, 118–132.

Pan, X., Wang, J., Zhang, X., Mei, Y., Shi, L., Zhong, G., 2018. A deep-learning model for the amplitude inversion of internal waves based on optical remote-sensing images. Int. J. Remote Sens. 39 (3), 607–618.

Pereira, P., Calliari, L., Holman, R., Holland, K., Guedes, R., Amorin, C., Cavalcanti, P., 2011. Video and field observations of wave attenuation in a muddy surf zone. Mar. Geol. 279 (1–4), 210–221.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al., 2019. Deep learning and process understanding for data-driven earth system science. Nature 566 (7743), 195.

Sakai, T., Battjes, J., 1980. Wave shoaling calculated from Cokelet's theory. Coastal Eng. 4, 65–84.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. ArXiv preprint arXiv:1801.04381.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. ArXiv preprint arXiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Stockdon, H.F., Holman, R.A., 2000. Estimation of wave phase speed and nearshore bathymetry from video imagery. J. Geophys. Res. Oceans 105 (C9), 22015–22033.

Stockdon, H.F., Holman, R.A., Howd, P.A., Sallenger Jr, A.H., 2006. Empirical parameterization of setup, swash, and runup. Coastal Eng. 53 (7), 573–588.

Stringari, C., Harris, D., Power, H., 2019. A novel machine learning algorithm for tracking remotely sensed waves in the surf zone. Coastal Eng.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI, vol. 4. p. 12.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proc. CVPR IEEE, pp. 2818–2826.

Thorne, P.D., Hurther, D., 2014. An overview on the use of backscattered sound for measuring suspended particle size and concentration profiles in non-cohesive inorganic sediment transport studies. Cont. Shelf Res. 73, 97–118.

Thornton, E.B., Guza, R., 1982. Energy saturation and phase speeds measured on a natural beach. J. Geophys. Res. Oceans 87 (C12), 9499–9508.