Remote Measurement of Active Whitecaps Using Deep Learning

XIN YANG,[a,b] HENRY POTTER,[a] SHUANG ZHANG,[a] AND MENG LYU[a]

[a] *Texas A&M University, College Station, Texas*
[b] *NOAA/OAR/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*

(Manuscript received 24 April 2024, in final form 31 January 2025, accepted 24 March 2025)

ABSTRACT: Whitecaps generated by wave breaking and air entrainment can be classified as active (stage *A*) or residual (stage *B*). Measurement of each stage individually is essential for accurate parameterization of air–sea interaction processes, but conventional methods used for separation in visible images are subjective. In this study, this problem is solved using a pipeline for active whitecap fraction measurement. In this pipeline, a new horizon detection method is developed to stabilize and rectify images, and a deep learning model based on U-Net is trained and validated to identify and extract active whitecaps. The model demonstrates robust prediction accuracy even when images are contaminated by sun glint. The model is applied to 48 h of video footage collected during a cruise in Gulf of Mexico. It is determined that, as a function of wind speed, the active whitecap fraction has significant variability and disparity compared to previous research. This finding indicates that secondary factors should be considered for accurate whitecap parameterization. This is explored using a random forest, which indicates that sea surface temperature, swell, and wave age are important to the active whitecap fraction. The precise impact of sea surface temperature is further explored using analyses of variance (ANOVA), which suggest it has a positive correlation with the active whitecap fraction.

KEYWORDS: Air-sea interaction; Remote sensing; Deep learning; Machine learning

1. Introduction

Whitecaps are generated during wave breaking when air is entrained into the surface ocean. Whitecaps can be observed when wind speeds are as low as 3 m s^{-1}, and it is estimated that whitecaps can cover up to 4% of the global ocean surface at any time (Blanchard 1983, 1963). Whitecap lifetime can be divided into two stages: the active stage (stage A) and the residual stage (stage B) (Monahan and Woolf 1989). At stage A, waves break under the continued influence of the wind. Large amounts of bubbles are generated at the wave crests and penetrate the water column by spilling and plunging. At stage B, bubbles rise to the surface and form patches of foam that linger behind the breaking crests. Whitecaps are often quantified by whitecap fraction W, which is the areal coverage of whitecaps on the sea surface. The total whitecap fraction is the sum of the active and residual whitecap fractions (i.e., $W = W_A + W_B$, where W_A and W_B are the active and residual whitecap fractions, respectively).

Whitecaps have a significant influence on physical and chemical processes at the air–sea interface. During stage A, wave breaking accounts for an important portion of wave energy dissipation. Hence, measurements of whitecaps, whitecap fraction, and breaking crest length are used to quantify energy dissipation (e.g., Gemmrich et al. 2008; Carini et al. 2015; Callaghan 2018). Also during stage A, wind tears breaking wave crests to generate spume droplets (e.g., Veron et al. 2012) which enhance air–sea gas transfer and heat flux (Andreas and Monahan 2000) and also influence tropical storm intensity (Andreas and Emanuel 2001). During stage B, the bubbles rise to the water surface and generate film and jet droplets through bubble bursting, which transform into sea salt aerosols (Deike 2022). Stage B whitecaps are especially important for ocean albedo because of their long lifetime (Frouin et al. 2001). Potter et al. (2015) showed that whereas active whitecaps have a lifetime range of 0.4–5 s, residual whitecaps can last tens of seconds. As such, stage B whitecaps have been found to contribute 1.5–40 times more to W than stage A (Callaghan et al. 2013). Whitecaps are also used to calculate air–sea transfer velocity (Asher et al. 2002), and they impact ocean acoustics, ocean mixing, surface roughness (Padmanabhan et al. 2006), as well as surface albedo (Koepke 1984). It is also essential to account for whitecaps in optical ocean color retrieval (Gordon and Wang 1994). Hence, robust understanding and accurate parameterization of whitecaps are critical for modeling these and many other processes at the air–sea interface that are associated with stage A or stage B whitecaps but not necessarily both.

Whitecap fraction has been measured extensively using visible images or videos collected from ships, offshore platforms, and aircraft (e.g., Monahan 1971; Wu 1988; Stramska and Petelski 2003; Mironov and Dulov 2008; Schwendeman and Thomson 2015b; Brumer et al. 2017; Yang and Potter 2021). Increased scattering of sunlight by whitecaps makes them brighter than the ambient water; therefore, whitecaps can be identified using image thresholding. Automatic thresholding determination algorithms have improved in recent years (e.g., Callaghan and White 2009; Bakhoday-Paskyabi et al. 2016). These algorithms increase image processing efficiency and mitigate uncertainty caused by manual threshold determination. However, extracting whitecaps from single images is still hindered by the critical problem of sun glint and uneven illumination (Callaghan and White 2009). Double-value thresholding methods are used to distinguish active whitecaps from

Corresponding authors: Xin Yang, xin.yang@noaa.gov; Meng Lyu, meng@tamu.edu

DOI: 10.1175/JTECH-D-24-0057.1

residual ones because active whitecaps tend to be brighter than residual foam (e.g., Hanson and Phillips 1999; Asher et al. 2002). However, separately identifying active and residual whitecaps is a great challenge due to the continuous and subtle change in brightness from stage A to stage B. Furthermore, the brightness change varies in space and time, rendering a single threshold inadequate.

Spaceborne remote sensing of the ocean surface brightness temperature T_B is used to measure whitecap fraction $W(T_B)$ on a global scale (e.g., Anguelova and Webster 2006). Separation between active and residual whitecaps based on T_B has also been explored (Potter et al. 2015; Salisbury et al. 2013). However, more data and analysis are required to identify stage A and B whitecaps from satellite-based T_B measurements (Potter et al. 2015). Uncertainties in modeling vertically stratified foam according to T_B sensitivity must be solved to provide a theoretical basis for whitecap discrimination using T_B (Anguelova and Gaiser 2011). Moreover, satellite-based T_B observations cannot identify individual wave breaking as the grid resolution is typically on the order of 10 km.

Infrared (IR) imagery provides dichotomic signals of active and residual whitecaps, which also have significant contrast against background water at IR wavelengths. Active and residual whitecaps are brighter (warmer) and darker (cooler), respectively, than background water (Marmorino and Smith 2005; Potter et al. 2015; Yang and Potter 2021). The clear dichotomic whitecap signals in IR imagery provide objective identification of different lifetime stages of whitecaps. The results from IR observations can be used to build more accurate parameterization of processes related to active and residual whitecaps. However, IR cameras with sufficient resolution and frame rate to capture the life stage transition are expensive and delicate, making the IR imagery method impractical for long-term and remote observations.

With the explosion of computer performance, machine learning, particularly deep learning, has gained popularity in many academic and industrial domains. The definition of machine learning is vague, but broadly speaking, algorithms that enable a computer to use data to "learn" and improve its performance on specific tasks can be called machine learning algorithms. Deep learning is a subset of machine learning that is often used to address computer vision problems, such as image recognition, object detection, object tracking, semantic segmentation, and image restoration (Voulodimos et al. 2018). Semantic segmentation involves dividing images into objects with semantic labels at the pixel level (Minaee et al. 2021). This technique can be applied to extract whitecaps from digital images. Wang et al. (2020) first developed a semantic segmentation model similar to a fully convolutional network (FCN) to extract whitecap fraction. Eadi Stringari et al. (2021) applied five mainstream deep neural networks to identify active whitecaps using ellipses and shared a labeled semantic segmentation dataset for active whitecaps with the scientific community. Sáez et al. (2021) trained and evaluated a deep learning model based on U-Net to identify coastal wave breaking. The rise of deep learning applications in whitecaps imagery data processing offers a promising method to enhance proficiency and

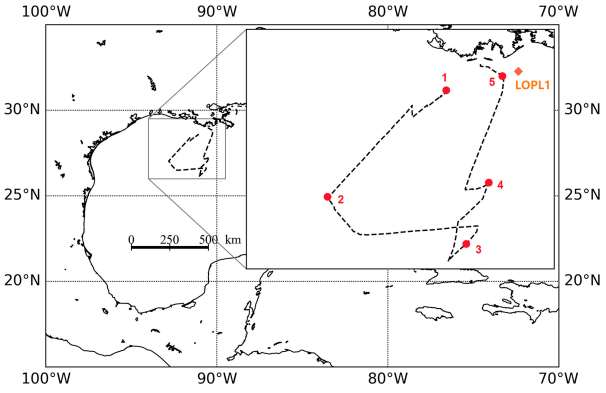

FIG. 1. Route of the cruise. Red dots denote the stations, and the orange diamond denotes the closest station with available simultaneous data.

achieve reliable performance. A key focus of this research is the development of accurate methods for separating W_A from W_B, which is essential for gaining a deeper understanding of the processes related to active and residual whitecaps.

Here, a pipeline based on deep learning for active whitecap extraction from images and W_A estimation is presented. This is a plug-and-play open-source model which includes a novel and robust horizon detection method to stabilize and rectify images collected by cameras at sea. The model is applied to video footage collected during a Gulf of Mexico cruise. Random forest (RF) regression is used to parameterize W_A based on environmental variables and to evaluate the importance of these variables in impacting W_A. Compared to a classical power-law parameterization using wind speed alone, the RF regression produces a more accurate model for predicting W_A. The layout of this paper is as follows: section 2 describes the data collection and processing, section 3 presents the preliminary analysis results, and section 4 summarizes the findings of this study.

2. Data and methods

a. *Instrumentation and data resources*

Data used in this paper come from the fieldwork described in Yang and Potter (2021) and will be briefly introduced here. The cruise, onboard the Research Vessel (R/V) *Pelican*, departed the Louisiana Universities Marine Consortium (LUMCON), Chauvin, Louisiana, on 4 March 2020, and returned on 8 March 2020. During the cruise, the ship stayed at five stations to collect data each day and transited between stations at night (Fig. 1). The wind speed measured onboard during the cruise was 4–18 m s^{-1}. The closest significant wave height measurements, made at station LOPL1, were 0.4–1.9 m according to the National Data Buoy Center (NDBC). Mean meteorological and oceanographic (met ocean) data collected onboard included air temperature (TA), sea surface temperature (SST), sea surface salinity (SSS), and wind speed U_{10}. These were processed and quality controlled by Lyu et al. (2021). Three GoPro Hero 8 Black digital video cameras were used to collect a total

of 60 h of video data. The field of view (FOV) of the lens in linear mode is 55.2° vertical FOV, 85.8° horizontal FOV, and the focal length is 19 mm. The linear mode FOV of GoPro camera corrects the lens distortion by using a transformation to map the distorted pixel locations back into a rectilinear projection and cropping the image to remove the edges most impacted by fisheye effect. The sampling rate was 60 Hz with 1920 × 1080 pixel resolution. One of the three cameras was mounted on the port side at a height of 4 m above the mean water level (MWL) and pointed toward the ocean surface with the horizon inside the frame at a tilt angle of 74°. The other two cameras were mounted on the port and starboard sides at a height of 7.6 m above the MWL with tilt angles of 73°. The data collected in deep water by these two cameras, which yielded about 48 h of video each, were processed for this study. The active whitecap fraction is determined by the average of 1-Hz images over 20 min (i.e., 1200 images) to decrease the uncertainty according to Callaghan and White (2009).

The fifth major global reanalysis produced by ECMWF (ERA5) provides simulated 1-hourly wave data on a global grid with 0.5° latitude–longitude grid cells across the globe. Based on previous studies (Sugihara et al. 2007; Callaghan et al. 2008; Brumer et al. 2017; Jia and Zhao 2019; Malila et al. 2022; Zhou et al. 2022), three variables are used to investigate their influence on active whitecap fraction: wave age (WG), significant height of total swell H_{swell}, and mean square slope of waves (MSQS). The H_{swell} and MSQS are from ERA5 simulation directly. WG is calculated using $gT/(2\pi \times u_*)$, where g is the acceleration due to gravity, T is the mean wave period, and u_* is the friction velocity. Mean wave period and friction velocity are also from ERA5 data. These three wave parameters were interpolated linearly to a 20-min resolution to align with W_A.

b. *Image rectification*

Measurement of whitecap fraction requires extraction of whitecaps from background water in digital images. To achieve this, the images must first be adjusted for the spatial pose of the camera (i.e., its position and orientation) due to the wave field which together mean each pixel accounts for a different, dynamic ocean surface area. Rectification is the projection of the digital image into Earth coordinates. Stabilization is removing the camera motion so that the pose of the camera is fixed for all images. There are two methods to measure the pose of the camera: one is using external equipment, such as an inertial motion unit, to determine the pose continuously, and the other is estimating the camera orientation based on images. For this project, we use the latter approach, building upon and improving a method by Schwendeman and Thomson (2015b). In their paper, a perspective transform equation is used for stabilization and rectification based on the angle of the horizon in each image and the pose of the camera when the ship is stable at port. As the only variable inferred from each image, an accurate estimate of the angle of the horizon determines the performance of this method. For this, Schwendeman and Thomson (2015b) used the Canny algorithm which detects connected edges in the image by determining the local maxima of brightness intensity. Ideally, for an image taken above the open sea, this will be the horizon because it marks the transition from the darker ocean surface and the brighter sky. Next, they used a Hough transform to extract straight lines in the images. Commonly, the horizon is the longest line in the image. However, a complicating factor for the performance of this method is poor illumination which can result in misidentification of the horizon or a horizon line that fails to stretch across the image. For example, in Fig. 2a, the brightness difference between the sky and ocean is small, and the ocean is very noisy (i.e., whitecaps and glint). For this image and many others like it, the Canny algorithm is unable to identify the horizon meaning stabilization and rectification cannot be completed. The Canny transform is typically unreliable when the sky is overcast and at high wind speed because there is more noise from the ocean surface.

Here, a new method to detect the horizon is presented which is more robust than that from Schwendeman and Thomson (2015b). The steps of the method are summarized below with details that follow.

1) A combination of binary and Otsu thresholding filters is applied to images to distinguish between the sky and ocean (Fig. 2b).
2) A dilation function is used twice, with different iterations, to eliminate noise in the image so that the black part denotes the sky, and the white part denotes the ocean (Fig. 2c).
3) Subtraction between the results of the dilation function yields the "horizon line" (Fig. 2d).
4) Because of the dilation process, the "horizon line" should be moved back several pixels depending on the number of dilation iterations so the resulting line is aligned with the horizon.

A line fitting function is used to yield the analytical formula of the straight line of the horizon (Fig. 2e). A Hough transform can also be used in place of the line fitting function for the same purpose.

Otsu thresholding identifies the maximum interclass variance as the best threshold to separate foreground and background under the assumption that they have different brightness intensity distributions. In this scenario, the sky and the ocean can be taken as background and foreground, respectively. An advantage of Otsu thresholding is that the threshold does not need to be determined subjectively. As shown in Fig. 2b, the pixels with intensity greater than the threshold are assigned black, and those with smaller intensity are assigned white. Some pixels in the sky and the ocean share the same brightness intensity, so it is impossible to identify them as the boundary using a single threshold. These pixels cause the noise in the image. The strategy here is to extract the pixels exactly at the boundary between the sky and the ocean and then take the line fitting result of these points as the approximation of the horizon. The dilation function, which is a convolution operation between the image and a kernel (a 3 × 3 square matrix is used in this project), is used to remove the noise. The bright regions will grow as more iterations of dilation are applied. In this project, the dilation

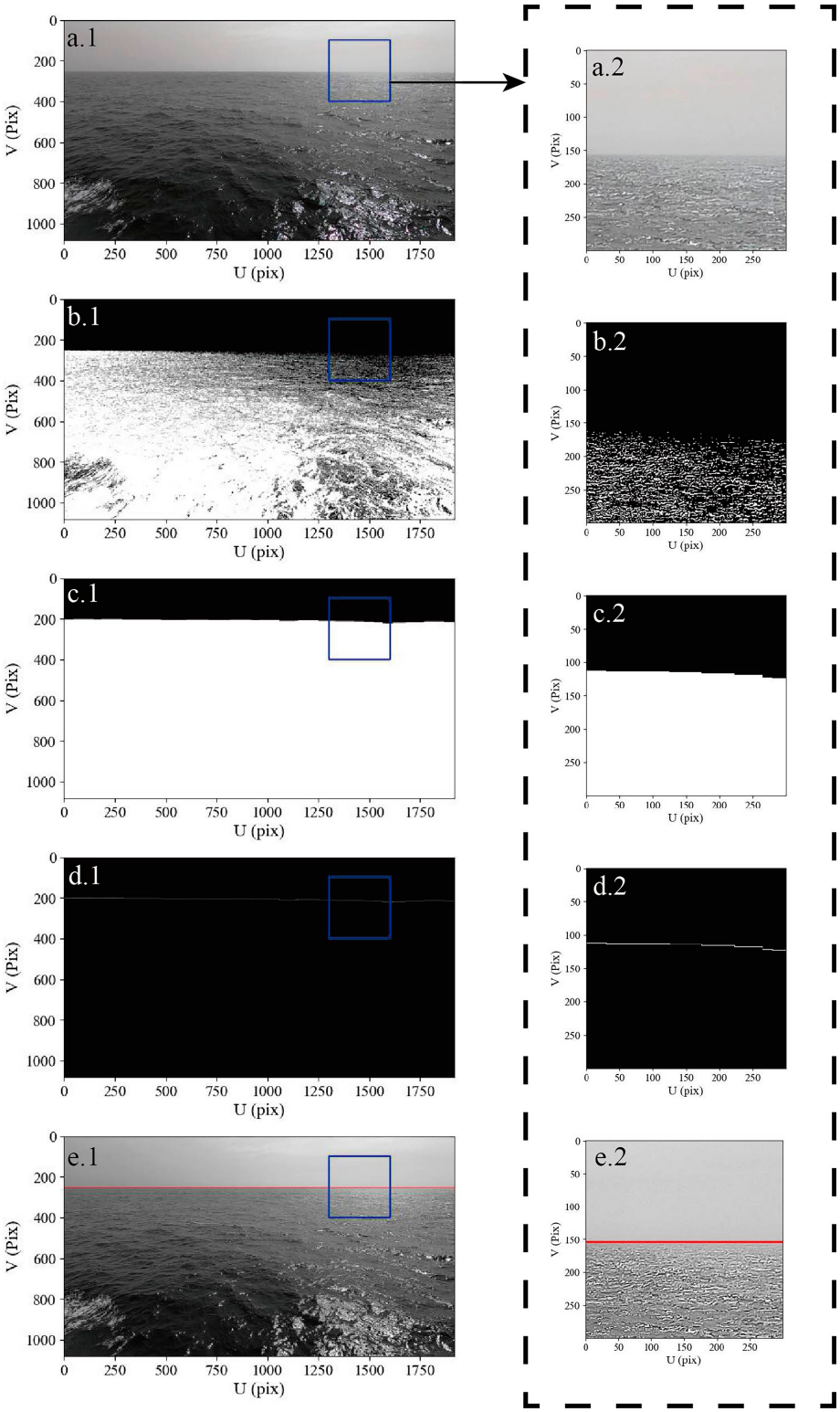

FIG. 2. Process of horizon detection. (a) Cropped original image, (b) image filtered by thresholding, (c) image processed by the dilation function, (d) subtraction between the results of the dilation function, and (e) detected horizon.

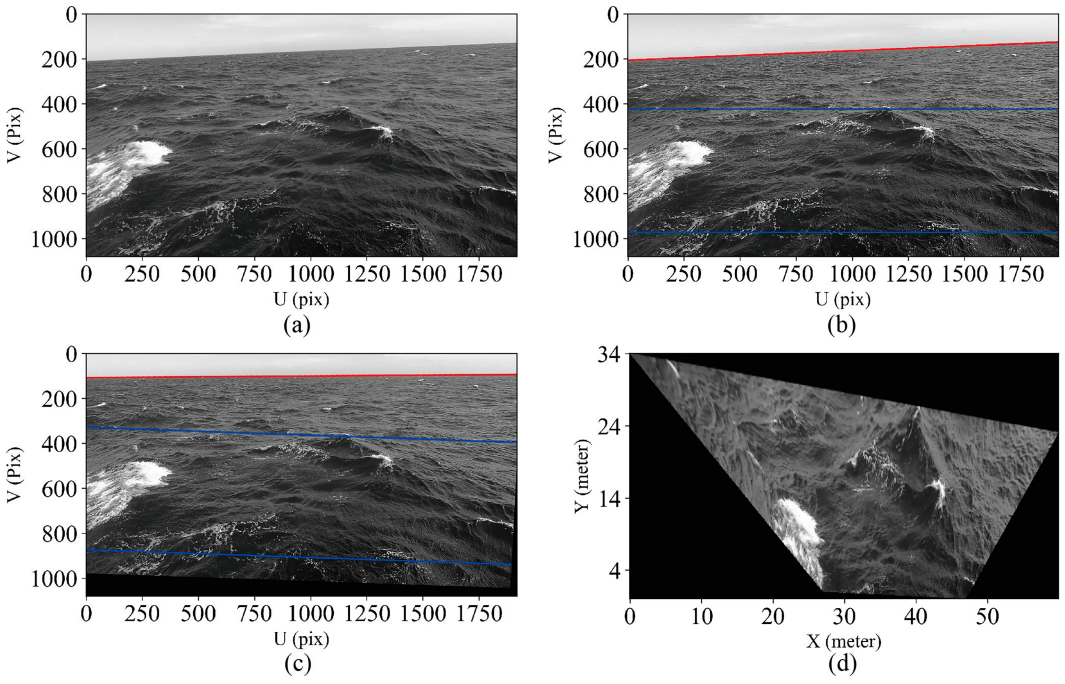

FIG. 3. Example of georectification. (a) The original image in grayscale; (b) the horizon is denoted by a red line, and the ROI is denoted by a blue rectangle; (c) image stabilization; and (d) image georectification results in a nadir view of the ROI.

function is applied to images twice while the numbers of iterations are set to 49 and 50. In both cases, the black pixels in the ocean area will be turned white (Fig. 2c). These two cases have one pixel difference at the boundary of the sky and ocean. Subtraction between these two cases results in a line with one pixel width. The resulting line should be moved down 50 pixels to cancel out the dilation to denote the actual horizon. The analytical expression of the horizon is determined by a least squares line fit.

The images can be stabilized using the angle of the detected horizon and the height of the camera above the mean water level, as shown in Figs. 3a–c. Furthermore, these parameters can also determine the simultaneous location and orientation of the camera (Schwendeman and Thomson 2015b), called extrinsic parameters. The intrinsic parameters of the camera can be calculated using the MATLAB built-in camera calibration toolbox. With the intrinsic and extrinsic parameters of the camera, the perspective transform equations can project the region of interest (ROI) in the image into the Earth coordinates (Fig. 3d). The perspective transform equations used in this paper are from Schwendeman and Thomson (2015a). The resulting images can then be used to calculate the whitecap fraction. The region of interest is chosen near the camera to avoid uncertainty caused by horizon detection according to Schwendeman and Thomson (2015b). Additionally, the pixels near the horizon represent much larger areas, which can result in greater uncertainty in whitecap fraction estimates. The bottom of the ROI is set away from the ship to avoid the influence of ship wakes while maintaining acceptable sea surface areas for analysis. Based on these considerations, the upper boundary of the ROI is set to one-quarter the distance from the horizon to the bottom of the image, and the lower boundary is set to one-tenth of the distance from the bottom of the image to the horizon, corresponding to approximately 9.36 m from the ship under steady conditions.

c. *Active whitecap fraction estimation*

Semantic segmentation is used to separate objects with semantic labels at the pixel level. Compared with traditional image segmentation methods, semantic segmentation based on deep learning has made remarkable improvements in accuracy (Minaee et al. 2021). In this study, the pixels representing active whitecaps need to be identified and extracted from each image. The first step to build a deep learning model is to choose a network backbone. Here, U-Net (Ronneberger et al. 2015), which was designed for biomedical image segmentation and won the International Symposium on Biomedical Imaging (ISBI) cell tracking challenge 2015, is used. U-Net, which combines a FCN and an encoder–decoder model, exhibits phenomenal performance even with a very small training data size. FCN uses deconvolution as up sampling to cast object recognition on each pixel and combines information of shallow and deep layers during up sampling to produce accurate segmentation (Long et al. 2015). An encoder–decoder model uses convolutional layers (encoder) to extract feature vectors and deconvolutional layers (decoder) to map feature vectors on segmentation masks (Minaee et al. 2021). The architecture of raw U-Net backbone has been modified to explore its potential (Siddique et al. 2020). For active whitecap extraction, the raw U-Net backbone, along with two variants whose

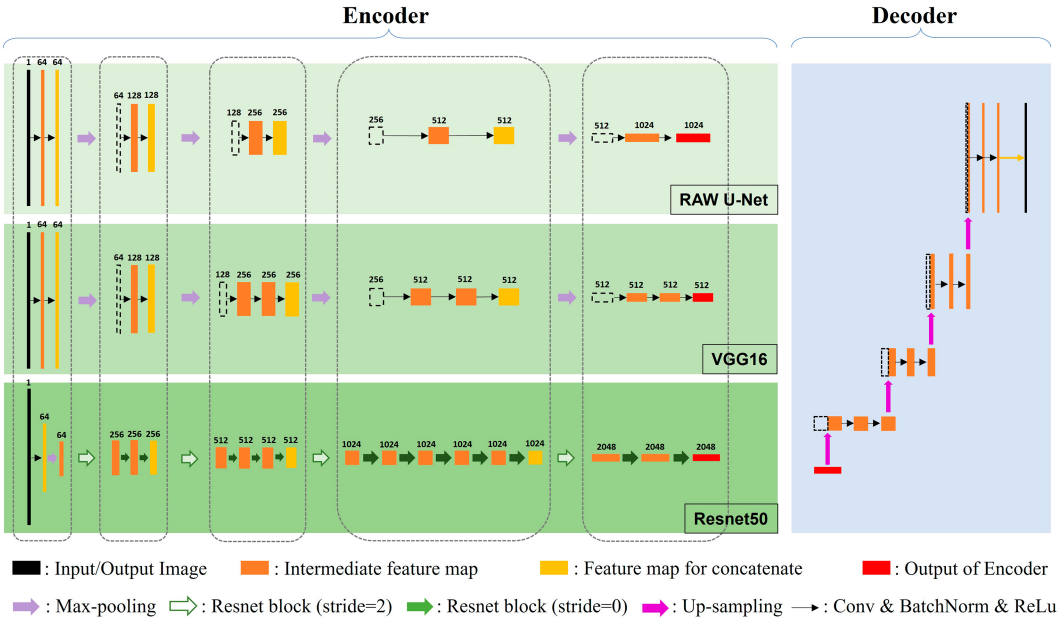

FIG. 4. Structure of U-Net, VGG16 variant, and ResNet50 variant. Each layer represents the feature map, and the number above the feature map is the number of channels. The black layers are input and output images which have 256 × 265 resolution. The orange layers are intermediate feature maps. The yellow layers are feature maps for crop and concatenate to corresponding decoder layers (dashed-lined layers in the decoder). The red layers are encoder outputs for up sampling. For more details about the ResNet block, see He et al. (2016).

encoder parts were replaced by VGG16 (Simonyan and Zisserman 2014) and residual neural network (ResNet50) (He et al. 2016), was trained (Fig. 4).

With a suitable model backbone, the second step is to choose a loss function to facilitate convergence of the deep learning model and evaluate the model performance using evaluation metrics. Active whitecaps extraction from digital images is a pixel-level binary classification problem with an extremely imbalanced dataset because the number of active whitecap pixels is much lower than that of the background. A confusion matrix (Table 1) is used to solve the active whitecap classification problem.

According to previous research (e.g., Anguelova and Webster 2006; Brumer et al. 2017), no more than 10% of the ocean surface is covered with whitecaps at low and moderate wind speeds, with active whitecaps just a fraction of this. Therefore, it is improper to use accuracy defined by (TP + TN)/(TP + FP+FN + TN) because the model can achieve a score over 90% even if it assigns all pixels as negative (i.e., no whitecaps). Therefore, precision, defined by TP/(TP + FP), and recall, defined by TP/(TP + FN), are more important than accuracy. In this study, F1 score, defined by the harmonic mean of precision and recall, is used to evaluate the performance of the model. The F1 score reaches 1 when the model makes perfect predictions and 0 when the model does nothing but make mistakes. A combination of binary cross entropy and Tversky loss is selected to be the loss function. Binary cross entropy is defined as

$$L_{\mathrm{BCE}}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]. \quad (1)$$

Here, y is the label, which is either 0 or 1; $\hat{y}$ is the prediction, which is a probability value; and N is the number of instances in a batch (Jadon 2020). Entropy, which was introduced by Shannon (1948), is used to evaluate the information content or uncertainty of a random variable. The idea of entropy is that the information content of an event is decided by the probability of the event. The information content is more abundant with smaller probability of occurrence. However, the probability distribution of a random variable is usually unknown in practice, so the entropy calculated for a random variable is cross entropy. The difference between cross entropy and entropy, which is proved to be nonnegative, is defined as Kullback–Leibler divergence, which comes to zero when the distribution of data equals the distribution of the population. Therefore, the minimum of cross entropy is a good strategy for the classifier. This idea is essentially the same as the maximum likelihood estimation.

Tversky loss is defined as

$$\mathrm{TL}(p, \hat{p}) = 1 - \frac{\mathrm{eps} + p\hat{p}}{\mathrm{eps} + p\hat{p} + \beta(1 - p)\hat{p} + (1 - \beta)p(1 - \hat{p})}. \quad (2)$$

TABLE 1. Confusion matrix.

		Prediction	
		Positive	Negative
Truth	Positive	TP (true positive)	FN (false negative)
	Negative	FP (false positive)	TN (true negative)

Here, p is the probability (or ratio) of (FP + TP), which means the probability that the prediction is 1 (or positive subjectively); $\hat{p}$ is the probability of (FN + TP), which means the probability that the label is 1 (or positive objectively); β is used to tune the weight of TN and FP; and eps is a number added to avoid zero division (Jadon 2020). Tversky loss can be turned into Dice loss when $\beta = 1/2$ (β is set to 1/2 in this project); therefore, it can be taken as a generalization of Dice loss, which is essentially 1 minus F1 score. The idea of Dice loss is to evaluate the ratio between the intersection of positive prediction and ground truth and the summation of positive prediction and ground truth. The lower the Dice loss, the better the prediction. Dice loss reaches zero and Dice score reaches one when the predicted segmentation equals the ground truth, or both are empty sets. In this study, the loss function used is simply the summation of binary cross entropy and Dice loss. This combination has good performance for data with imbalanced classes (Yeung et al. 2022). Through the above procedure, the classification problem is turned into the convex optimization problem, where the objective is to minimize the loss function. The strategy is to update the parameters in the direction of the gradient descent of the convex optimization to approach the loss function minimum. This process can be described as

$$\mathbf{w} = \mathbf{w} - \frac{\theta}{B}\sum_{i\in B}\frac{\partial l^{(i)}(\mathbf{w})}{\partial \mathbf{w}}, \tag{3}$$

where $\mathbf{w}$ is the parameters of the model, θ is the learning rate, B is the batch size, and l is the loss function. Gradient descent is to make the variable decrease with a fixed step size in the direction of the gradient of this function. Different gradient descent methods have been widely used, such as stochastic gradient descent (SGD), minibatch gradient descent, SGD with momentum, adaptive gradient descent (AdaGrad), root-mean-square propagation (RMS-Prop), AdaDelta, adaptive moment estimation (Adam), and Rectified Adam (RAdam). Here, RAdam is implemented in the training processes because it improves training stability and efficiency. The learning rate will be scaled by 0.2 when the metrics for the model stop decreasing for two epochs, and it will wait for two more epochs before resuming the gradient descent process. This happens automatically (Fig. 5b).

The training and validation dataset used in this project is from Eadi Stringari et al. (2021). There are 1553 samples in the training dataset and 366 samples in the validation dataset. Some examples of data and mask pairs are shown in Fig. 5a. The active whitecaps in the image are labeled as 1, and the rest of the pixels are labeled as 0. A confusion matrix normalized by each true label, as shown in Fig. 5c, was plotted for the validation dataset.

During the training iterations (or epochs), the training and validation loss along with model metrics are calculated for each backbone to assess model performances, as shown in Fig. 6. All three models are good at identifying pixels for active whitecaps and reach their best metrics around 40 epochs. The validation loss reaches a plateau with increasing epochs. However, ResNet50 exhibits a lower loss value of ~0.02 when compared to VGG16 and raw U-Net, in addition to having a higher F1 score. These three models have similar performance in precision (Fig. 6b), but ResNet50 does better in recall (Fig. 6c). This indicates that ResNet50 has better control in TN, which is not obvious in the confusion matrix due to the large number of negative samples. The F1 scores of VGG16 and raw U-Net reach around 0.78 and that of ResNet50 can be pushed to around 0.81, meaning these models are very good at active whitecap extraction. For comparison, a conventional thresholding method developed by Kleiss and Melville (2011) was built and evaluated using the same dataset, resulting in an F1 score of approximately 0.28.

The best trained model based on ResNet50 was applied to GoPro images collected during the Gulf of Mexico cruise. Several examples of results are shown in Fig. 7. The entire active whitecaps extraction pipeline completes rectification and W_A estimation automatically. The model performs well, as evidenced by the F1 score of 0.81, the loss of 0.25, the precision of 0.86, and the recall of 0.8, even though the processed data and the training data were collected with different instruments and under different conditions.

To verify the model's sun-glint removal capability, the whitecap images were manually inspected, and those with sun glint were removed. Then, datasets with and without contaminated images were fed into the model to determine W_A. Each resulting W_A is an average of 1-Hz images over a 20-min observation (i.e., 1200 frames). For glint-free runs, the number of frames was reduced slightly. The results in Fig. 8 show that the model is very good at identifying sun glint and eliminating it from estimates of W_A. The ability of the model to remove the disturbance of sun glint is unprecedented compared to traditional methods. The code for training and implementation is available at GitHub (https://github.com/yangxin1022/active_whitecap_DL.git).

d. RF regression

Random forest (RF) regression is a machine learning algorithm that combines multiple decision tree models to improve the robustness of regression. It can effectively handle large datasets with many input features and complex interactions between them. RF regression offers a feature importance metric that can be used to rank the relevance of input variables for predicting the target variable. The feature importance is calculated by measuring the increase in mean squared error when a feature variable is permuted across the observations (permutation test). The idea is that if a feature is useful for prediction, then the inclusion of that feature in the RF model should lead to a significant reduction in mean squared error. As such, when we introduce random permutations to that specific feature and input this modified feature, along with the remaining features, into the trained machine learning (ML) model, we anticipate a consequent increase in the mean squared error. In this study, we use RF regression to predict W_A using multiple environmental factors and permutation feature importance to evaluate the influence of these factors.

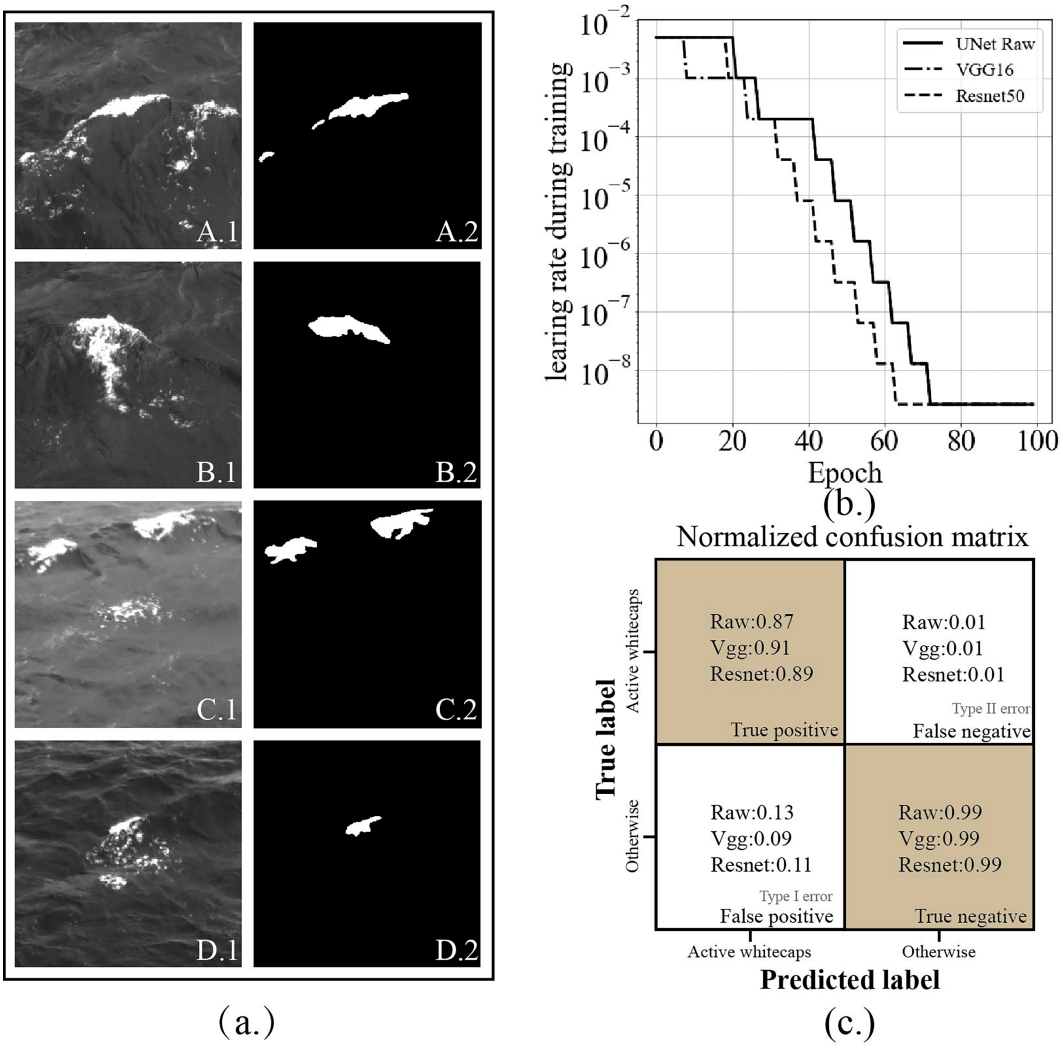

FIG. 5. (a) Examples of the training and validation dataset. (left) The original images and (right) masks. The white pixels denote active whitecaps. (b) Scheduled learning rate change during training. (c) Normalized confusion matrix for the best-validated models based on raw U-Net, VGG16, and ResNet50. Each row of the matrix is normalized by a corresponding true label.

3. Results and discussion

a. W_A versus wind speed

Wind speed U_{10} is typically considered the main driver of whitecap formation, and empirical expressions of whitecap fraction are often formulated as a function of U_{10} (Anguelova and Webster 2006; Brumer et al. 2017), though great variability remains. Figure 9 presents simultaneous wind speed and W_A collected on each day during the cruise. The average of W_A collected from both the port and starboard sides is used to determine each point. On 4 March (Fig. 9a), the wind speed ranges from 3.9 to 5.1 m s^{-1}, and W_A is low, around 0.02%–0.05%. On 5 March (Fig. 9b), the wind speed increased abruptly to around 18 m s^{-1}, and W_A exhibits a strong correlation with wind speed and reaches a maximum value of 0.17%. On 6 March (Fig. 9c), the wind speed stays between 13.6 and 16 m s^{-1}, while W_A shows large fluctuations, ranging from 0.08% to 0.42%. After 1000 UTC, wind speed begins to decrease, which is followed by a decrease in W_A with a lag of about 1 h. On 7 March (Fig. 9d), the wind speed keeps decreasing, while W_A oscillates between 0.05% and 0.15%. The observed variability in W_A at specific wind speeds suggests that secondary factors may also contribute to active whitecap area.

Following the mainstream approach, W_A as a function of U_{10} is parameterized. The threshold-power-law (e.g., Schwendeman and Thomson 2015a) fit is applied to the data. The polynomial regression is solved using the Levenberg–Marquardt algorithm by minimizing the sum of squares of the standard residuals. The formula is

$$W_A = 3.64 \times 10^{-5} \times (U_{10} + 1.59)^{1.30},\ 4 < U_{10} < 18\,\mathrm{m\,s^{-1}}. \quad (4)$$

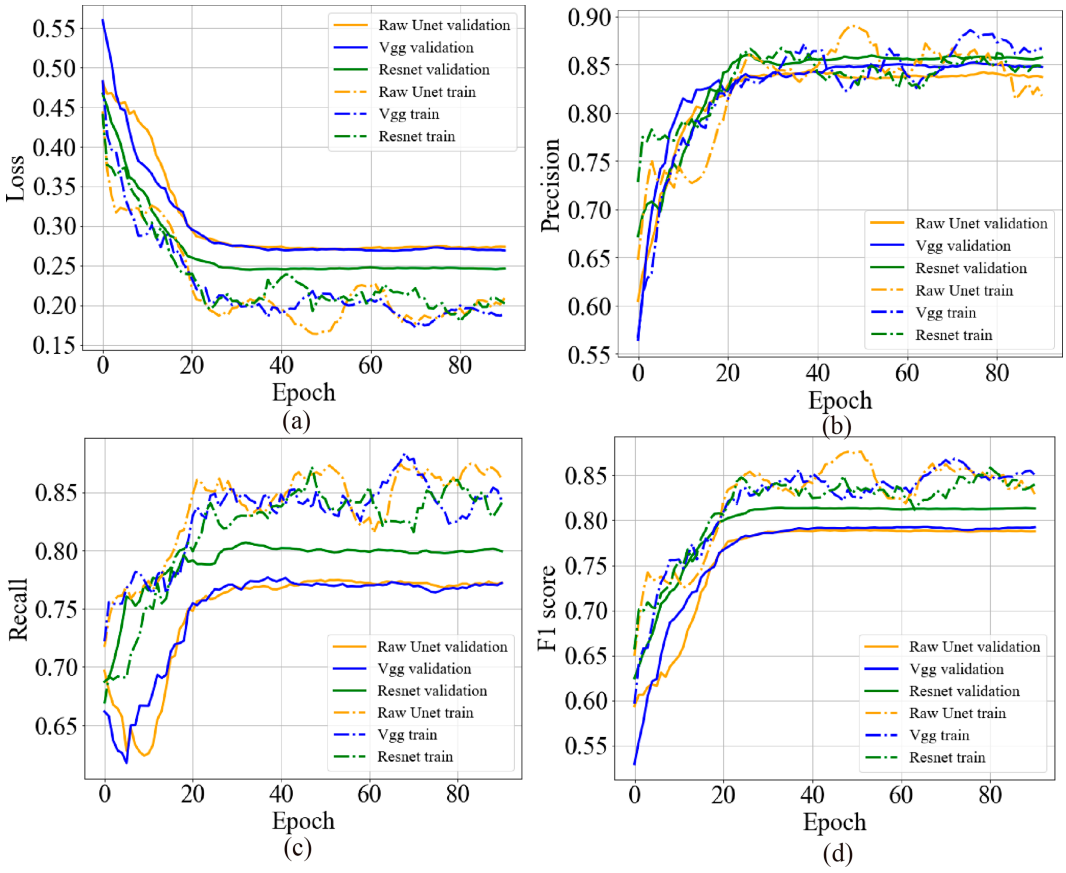

FIG. 6. (a) Loss function value, (b) precision, (c) recall, (d) F1 score for the training and validation data. The plots are smoothed by a 10-epoch running mean.

This equation is derived from observational data with U_{10} ranging from 4 to 18 m s^{-1}. The results for all 20-min-averaged W_A measurement, as well as previous parameterization results, are shown in Fig. 10. The coefficient of determination R^2 = 0.44. This parameterization exhibits larger W_A values than previously reported at lower wind speed ($U_{\text{wind}} < 10$ m s^{-1}) and smaller values at higher wind speed ($U_{10} > 13$ m s^{-1}). Hence, our result shows W_A has less wind speed dependence than previous studies except Salisbury et al. (2013) who used data collected by satellite-based microwave remote sensing data and found a similar pattern, though at a higher magnitude.

The traditional cube law for total whitecap fraction parameterization is also applied to build the relationship between W_A and U_{10}. The cube law shows good approximation at higher wind speed ($U_{10} > 13$ m s^{-1}), but underestimates W_A at lower wind speed ($U_{10} < 10$ m s^{-1}). The coefficient of determination is 0.14. The result is

$$W_A = 4.11 \times 10^{-7} \times U_{10}^3,\ 4 < U_{10} < 18\,\text{m s}^{-1}. \quad (5)$$

The poor performance of the cube-law fit indicates that the relationship between active whitecaps and wind speed is very different from that between total whitecaps and wind speed. Great deviation between the data and parameterizations from previous research indicates that W_A is impacted by local environmental factors and that W_A cannot be predicted by wind speed alone.

There is no agreement about the relationship between energy dissipation rate and either W or W_A. Many previous studies show a linear relationship between W and energy dissipation rate (Hwang and Sletten 2008; Scanlon et al. 2016), while some others suggest that W_A presents a more significant linear relationship with energy dissipation rate (Anguelova and Hwang 2016; Kraan et al. 1996; Scanlon et al. 2016). The exponent of 1.3 in Eq. (4) suggests that W_A increases more rapidly as wind speed increases. Based on the assumption that W_A has a linear relationship with energy dissipation rate, the exponent in $W_A(U_{10})$ formula indicates that wave energy dissipation can be intensified by increasing wind speed, and a mature wave field may have different relationships with a growing sea. However, the parameterizations of W dependent on wind speed from previous literature usually have an exponent around 3, as shown in Eq. (5) (Anguelova and Webster 2006; Brumer et al. 2017). It indicates that W can be more sensitive to the change of wind speed, and the linear relationship between energy dissipation rate and W or W_A cannot coexist. More comprehensive data are required to achieve a deeper understanding.

Almost every campaign for whitecap imagery data collection developed and applied its own data processing techniques. The development of these methods involves a degree of

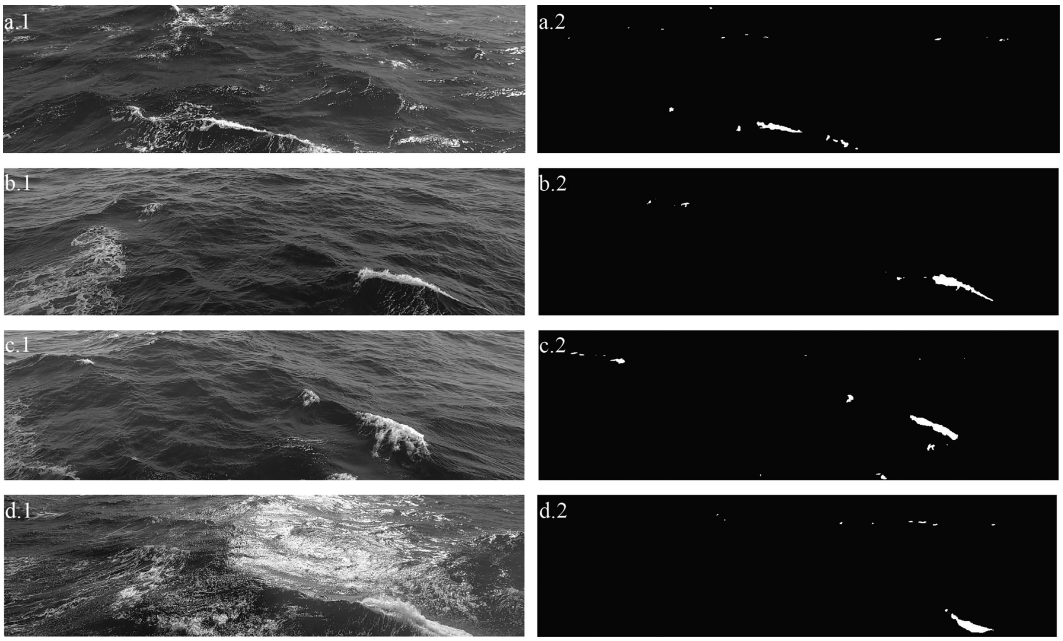

FIG. 7. Examples of images (without rectification) processed by the ResNet50 variant model. (left) The original images. (right) Masks with white pixels denoting active whitecaps.

subjectivity, and the results have not been validated or evaluated quantitatively. While the thresholding-based methods varied in some details, they shared a common data processing philosophy, potentially leading to similar uncertainties across studies. The deep learning method proposed in our study is designed to be objective and entirely data driven. We have made both the code and model open access, encouraging fine-tuning and optimization.

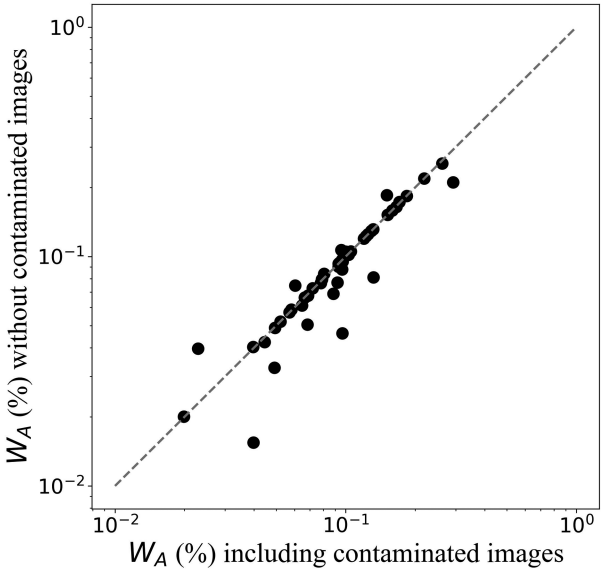

FIG. 8. Comparison between 20-min-averaged W_A with and without images contaminated by glint.

b. *Analysis of multiple factors using RF regression*

Results from the prior section support the idea that wind speed alone is not enough to parameterize active whitecap fraction. This means that other environmental factors need to be considered. Many studies have demonstrated the influence of sea state on whitecap fraction. Brumer et al. (2017) shows that incorporating wave variables can decrease the variability of wind-speed-only parameterizations. Jia and Zhao (2019) state that swell can depress whitecap fraction. Malila et al. (2022) report that wave group passage can enhance active whitecap fraction by fivefold. Zhou et al. (2022) use deep neural networks to construct a multivariate whitecap fraction parameterization and indicate that significant wave height is a secondary factor. In this study, SST, TA, WG, H_{swell}, and MSQS were included in the analysis along with U_{10}.

Here, we use RF regression to predict W_A using H_{swell}, SST, TA, WG, MSQS, and U_{10}, and also use feature importance analysis to assess the extent to which these factors drive the change of W_A in the regression process. Cross validation is used to train and evaluate the model to mitigate the risk of overfitting or underfitting. A grid search over a range of values coupled with cross validation was conducted to find the optimal set of hyperparameters based on the best metrics (minimal residual mean squared error) obtained.

The random forest regression result of W_A is compared to the measured W_A, as shown in Fig. 11a. The mean coefficient of determination from cross validation is 0.70, which is improved by 0.26 compared to the curve fitting in section 3a. Note that this improvement is also a conservative estimate as testing data were used to judge the model performance, while in the curve fitting, all data are utilized. The RF regression

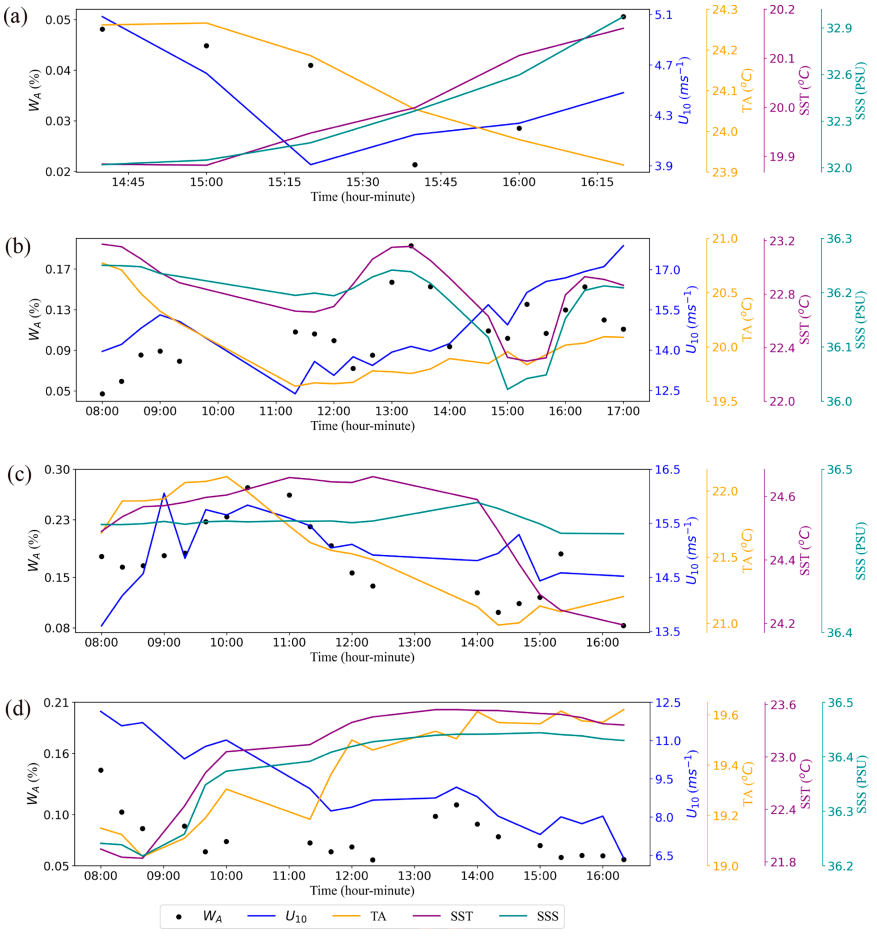

FIG. 9. Twenty-minute-averaged met-ocean data and W_A during the Gulf of Mexico cruise. (a) Data from 4 Mar, (b) data from 5 Mar, (c) data from 6 Mar, and (d) data from 7 Mar. Note, the x axis in each subplot covers a different length of time.

provides an accurate and robust parameterization of W_A. It shows that H_{swell}, SST, and WG are as important in predicting W_A as U_{10}, while TA and MSQS have limited influence (Fig. 11b). The swell conditions during the cruise were analyzed using the ERA5 dataset, following the method from Sugihara et al. (2007). Swell was categorized as 1) following swell, when the swell direction is within ±45° of the wind direction; 2) opposing swell, when the swell direction is within ±45° of

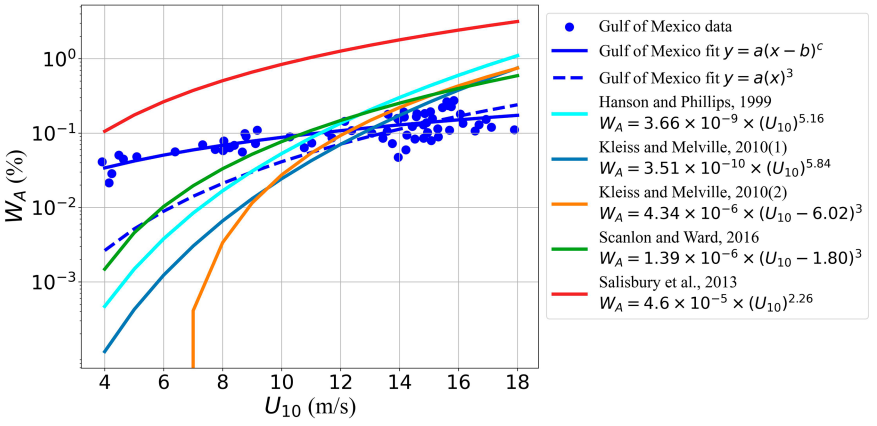

FIG. 10. Active whitecap fraction W_A as a function of U_{10}. The dots show the 20-min-averaged W_A computed from the Gulf of Mexico data. The lines are parameterizations based on the data in this study and from previous whitecap research.

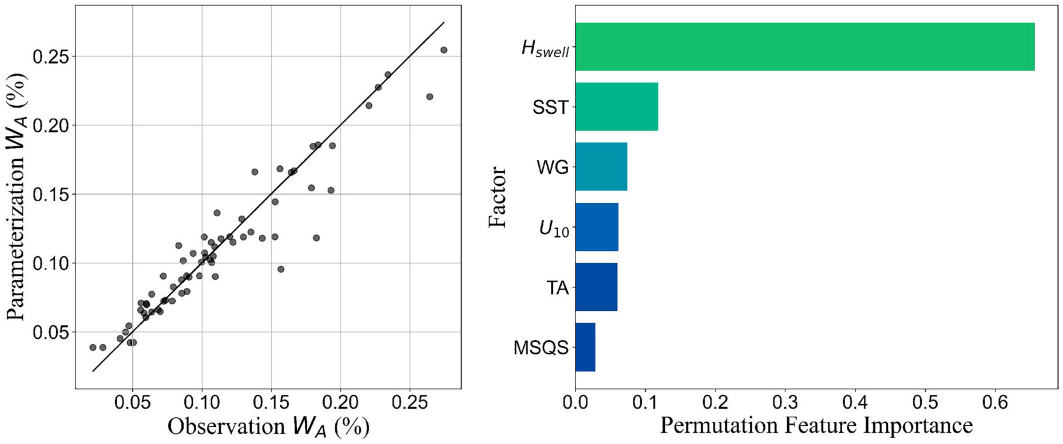

FIG. 11. (a) The parameterization results for W_A based on RF against the observational data. The straight line, $y = x$, is plotted for comparison; (b) the feature importance for the variables used in the RF.

the opposite wind direction; and 3) cross swell, for all other cases. The results revealed a diverse pattern of swell conditions throughout the cruise period. Data collected on 5 March were characterized by opposing swell conditions, while measurements taken around 1600 UTC 4 March and 7 March exhibited cross swell conditions. The remaining data were predominantly obtained under the following swell conditions.

c. *Secondary influence of SST*

The RF regression provides a qualitative evaluation of the importance of certain environmental factors in W_A. Given that the wave data are derived from numerical models at relatively coarse spatial resolution and have undergone temporal interpolation, they are more appropriate for bulk assessments than for quantitative analysis. As such, further investigation neglects wave variables and instead focuses on met-ocean parameters which were recorded in situ and at high resolution. To investigate the relationship between SST and W_A, it is necessary to remove the strong influence of wind speed, which is done by dividing the dataset into three scales (scale 1–3): 6 m s^{-1} $< U_{10} \leq$ 10 m s^{-1}, 10 m s^{-1} $< U_{10} \leq$ 14 m s^{-1}, and 14 m s^{-1} $< U_{10}$. This ensures that the influence of U_{10} is relatively small on each scale. The data with U_{10} less than 6 m s^{-1} are excluded due to their limited sample size (i.e., only six points). The dataset is then separated into two groups by the median SST within the wind speed scale. A Student's t test was conducted on U_{10} for each group, resulting in no statistical difference between them (i.e., p values are 0.45, 0.14, and 0.23, respectively). One-way analysis of variance (ANOVA) is then performed to determine the influence of SST on W_A for each wind speed scale. In scale 2, there are 16 samples, and there is no statistical difference between the averages of the groups, as shown in Fig. 12b(1). In scales 1 and 3, there are 13 and 32 samples, and the p values are 0.03 and 0.0035, respectively. This means, at a very high level of confidence, SST impacts W_A within these wind speed bins. The results in Fig. 12a(1) and 12c(1) agree with the analysis in section 3b. However, it remains unclear why the influence of SST is insignificant in scale 2. It is possible that the sample size in moderate and low U_{10} is too small to observe a statistically significant impact of SST (i.e., not enough statistical power to detect existing differences among the groups).

Next, the data are divided into groups according to the scales of U_{10} and SST, and a two-way ANOVA is applied. To ensure the sample size of each factor is as similar as possible, U_{10} is divided into three scales: $U_{10} \leq$ 10 m s^{-1}, 10 m s^{-1} $< U_{10} \leq$ 14 m s^{-1}, and $U_{10} >$ 14 m s^{-1}, while SST is divided into two scales: SST $\leq$ 23°C and SST $>$ 23°C. For this ANOVA (figure not shown), the p values for U_{10} and SST group are 0.0007 and 0.03, respectively, while that for the interaction between the two factors is 0.80. This means U_{10} and SST both have significant influence on W_A, respectively (as previously established), while the effect of SST is not dependent on U_{10} (or vice versa). Importantly, this means that the relationship between SST and W_A is not a spurious correlation. Hence, these results are consistent with Callaghan et al. (2014), who found that air entrainment increased with warmer water in a laboratory experiment leading to increased active wave breaking area. A positive correlation between SST and W was also found by Monahan and O'Muircheartaigh (1986) and by Albert et al. (2016), who suggested SST should be included in W parameterization. Hooker et al. (2020) analyzed 17 whitecap datasets and found the positive relationship between SST and the exponent of power-law fitting of $W_B(U_{10})$. Furthermore, Liu and Yang (2022) showed that parameterization with positive SST dependence fitted observational data better than that with negative SST dependence. They also proposed a parameterization incorporating both U_{10} and SST, which yielded a more accurate fit for W.

4. Summary

The evolution of whitecaps formed by wave breaking can be divided into stage A (active) and stage B (residual)

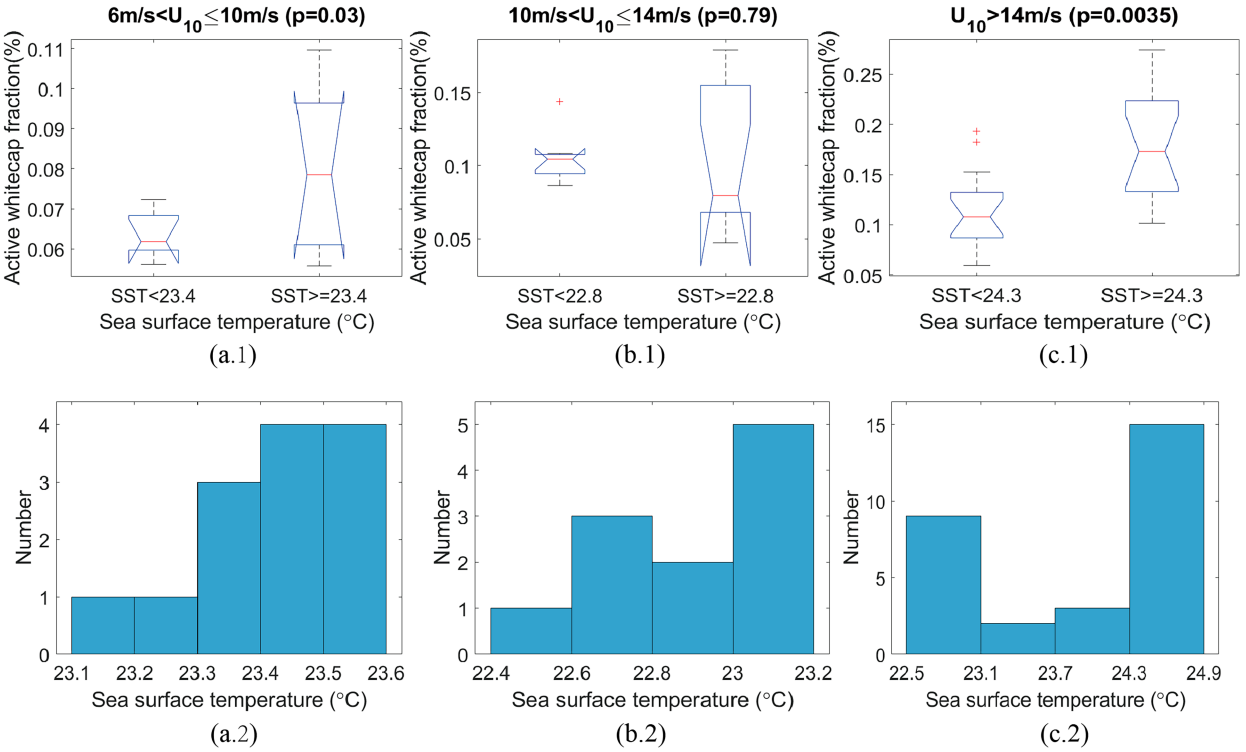

FIG. 12. (a1)–(c1) Boxplots for W_A characterized by SST in each wind speed scale. Central red lines are the median; box edges are the first and third quartiles. The whiskers extend to the limits for data points that are not considered outliers, which are denoted by red plus signs. The p values for (a1)–(c1) are 0.03, 0.79, and 0.0035, respectively. (a2)–(c2) Histograms for SST in each wind speed scale.

according to their prominently different characteristics. Active whitecaps accompany air entrainment into the water during wave breaking, while residual whitecaps are patches of bubbles lingering on the water surface. Many air–sea interaction processes are related to active and residual whitecaps but not necessarily both. Therefore, it is crucial to measure active and residual whitecap fractions separately. While whitecaps have been measured extensively, few studies have focused on individual stages. This is largely because the extraction of whitecap stages from visible imagery is challenging because the transition from active to residual stages is subtle. Furthermore, features of active whitecaps are either difficult to capture or impractical to extract using automatic programming.

This study presents a pipeline to quantify the active whitecap fraction through the analysis of visible imagery. Data were collected during the same Gulf of Mexico cruise used in Yang and Potter (2021). The algorithm introduced by Schwendeman and Thomson (2015b) was used to stabilize and rectify images from shipboard cameras, while a new horizon detection method was developed to enhance the robustness. The new method incorporates a thresholding filter, dilation iteration, and linear regression to achieve great performance even in challenging conditions such as poor illumination and contamination from sun glint. A U-Net-based deep learning model is trained and validated using a dataset from Eadi Stringari et al. (2021) to extract the pixels of active whitecaps from visible images automatically. The model has a high level of accuracy with an F1 score on the validation data of 0.81. Until now, dealing with sun glint and sky reflection contamination has been a great challenge when extracting whitecap fraction from visible images. However, this deep learning model has phenomenal performance when identifying active whitecaps in contaminated images. The code for the entire data processing pipeline and the trained model can be accessed on GitHub and is made freely available to the community.

This study used the aforementioned whitecap deep learning model to determine W_A from 48-h video captured using GoPro cameras. The wind speed–dependent relationship of active whitecap fraction is investigated. The power law is used to parameterize W_A as a function of U_{10}. The best fit was found to be $W_A = 3.64 \times 10^{-5} \times (U_{10} + 1.59)^{1.30}$. For any given wind speed, there is significant scatter in W_A, which indicates that secondary factors need to be considered. The random forest regression was employed to parameterize W_A, resulting in $R^2 = 0.07$ on unseen observations, which showed a better fit to the data compared to the power-law parameterization ($R^2 = 0.44$) using wind speed alone. Additionally, the feature importance analysis revealed that H_{swell}, WG, and SST are crucial components in the parameterization process. One-way and two-way ANOVA are applied to W_A to investigate the secondary influences of SST. The results indicate that SST has a significant impact on W_A. To ensure a more

comprehensive and robust investigation, additional in situ data are needed in future studies.

Acknowledgments. This research was funded by the National Science Foundation under Grant 1829986. Additional funding was provided by the Texas A&M Department of Oceanography to help support graduate and undergraduate students who participated in the research cruises. We are very grateful to Dr. Geoff Smith of the Naval Research Laboratory Remote Sensing Division who provided equipment and operational support during both cruises as well as postprocessing expertise. We thank the captain and crew of the R/V *Pelican* for two successful cruises. Finally, we express our appreciation to the dozen graduate and undergraduate students who all contributed to success of the research cruises.

Data availability statement. Cruise data are available through the Rolling Deck to Repository (R2R) at https://www.rvdata.us for cruises PE20-08 and PE20-17. Raw video footage is available upon request by contacting the corresponding author. The reference to the training and validation dataset is cited in the paper and listed in the reference section. The code for training and implementation is available at GitHub: https://github.com/yangxin1022/active_whitecap_DL.git.

REFERENCES

Albert, M. F. M. A., M. D. Anguelova, A. M. M. Manders, M. Schaap, and G. de Leeuw, 2016: Parameterization of oceanic whitecap fraction based on satellite observations. *Atmos. Chem. Phys.*, **16**, 13 725–13 751, https://doi.org/10.5194/acp-16-13725-2016.

Andreas, E. L., and E. C. Monahan, 2000: The role of whitecap bubbles in air–sea heat and moisture exchange. *J. Phys. Oceanogr.*, **30**, 433–442, https://doi.org/10.1175/1520-0485(2000)030<0433:TROWBI>2.0.CO;2.

——, and K. A. Emanuel, 2001: Effects of sea spray on tropical cyclone intensity. *J. Atmos. Sci.*, **58**, 3741–3751, https://doi.org/10.1175/1520-0469(2001)058<3741:EOSSOT>2.0.CO;2.

Anguelova, M. D., and F. Webster, 2006: Whitecap coverage from satellite measurements: A first step toward modeling the variability of oceanic whitecaps. *J. Geophys. Res.*, **111**, C03017, https://doi.org/10.1029/2005JC003158.

——, and P. W. Gaiser, 2011: Skin depth at microwave frequencies of sea foam layers with vertical profile of void fraction. *J. Geophys. Res.*, **116**, C11002, https://doi.org/10.1029/2011JC007372.

——, and P. A. Hwang, 2016: Using energy dissipation rate to obtain active whitecap fraction. *J. Phys. Oceanogr.*, **46**, 461–481, https://doi.org/10.1175/JPO-D-15-0069.1.

Asher, W., J. Edson, W. McGillis, R. Wanninkhof, D. T. Ho, and T. Litchendorf, 2002: Fractional area whitecap coverage and air-sea gas transfer velocities measured during GasEx-98. *Gas Transfer at Water Surfaces*, *Geophys. Monogr.*, Vol. **127**, Amer. Geophys. Union, 199–203, https://doi.org/10.1029/GM127p0199.

Bakhoday-Paskyabi, M., J. Reuder, and M. Flügge, 2016: Automated measurements of whitecaps on the ocean surface from a buoy-mounted camera. *Methods Oceanogr.*, **17**, 14–31, https://doi.org/10.1016/j.mio.2016.05.002.

Blanchard, D. C., 1963: The electrification of the atmosphere by particles from bubbles in the sea. *Prog. Oceanogr.*, **1**, 73–202, https://doi.org/10.1016/0079-6611(63)90004-1.

——, 1983: The production, distribution, and bacterial enrichment of the sea-salt aerosol. *Air-Sea Exchange of Gases and Particles*, P. S. Liss and W. G. N. Slinn, Eds., ATO ASI Series, Vol. 108, Springer, 407–454, https://doi.org/10.1007/978-94-009-7169-1_7.

Brumer, S. E., C. J. Zappa, I. M. Brooks, H. Tamura, S. M. Brown, B. W. Blomquist, C. W. Fairall, and A. Cifuentes-Lorenzen, 2017: Whitecap coverage dependence on wind and wave statistics as observed during SO GasEx and HiWinGS. *J. Phys. Oceanogr.*, **47**, 2211–2235, https://doi.org/10.1175/JPO-D-17-0005.1.

Callaghan, A., G. De Leeuw, L. Cohen, and C. D. O'Dowd, 2008: Relationship of oceanic whitecap coverage to wind speed and wind history. *Geophys. Res. Lett.*, **35**, L23609, https://doi.org/10.1029/2008GL036165.

Callaghan, A. H., 2018: On the relationship between the energy dissipation rate of surface-breaking waves and oceanic whitecap coverage. *J. Phys. Oceanogr.*, **48**, 2609–2626, https://doi.org/10.1175/JPO-D-17-0124.1.

——, and M. White, 2009: Automated processing of sea surface images for the determination of whitecap coverage. *J. Atmos. Oceanic Technol.*, **26**, 383–394, https://doi.org/10.1175/2008JTECHO634.1.

——, G. B. Deane, and M. D. Stokes, 2013: Two regimes of laboratory whitecap foam decay: Bubble-plume controlled and surfactant stabilized. *J. Phys. Oceanogr.*, **43**, 1114–1126, https://doi.org/10.1175/JPO-D-12-0148.1.

——, M. D. Stokes, and G. B. Deane, 2014: The effect of water temperature on air entrainment, bubble plumes, and surface foam in a laboratory breaking-wave analog. *J. Geophys. Res. Oceans*, **119**, 7463–7482, https://doi.org/10.1002/2014JC010351.

Carini, R. J., C. C. Chickadel, A. T. Jessup, and J. Thomson, 2015: Estimating wave energy dissipation in the surf zone using thermal infrared imagery. *J. Geophys. Res. Oceans*, **120**, 3937–3957, https://doi.org/10.1002/2014JC010561.

Deike, L., 2022: Mass transfer at the ocean–atmosphere interface: The role of wave breaking, droplets, and bubbles. *Annu. Rev. Fluid Mech.*, **54**, 191–224, https://doi.org/10.1146/annurev-fluid-030121-014132.

Eadi Stringari, C., P. Veras Guimarães, J. F. Filipot, F. Leckler, and R. Duarte, 2021: Deep neural networks for active wave breaking classification. *Sci. Rep.*, **11**, 3604–3612, https://doi.org/10.1038/s41598-021-83188-y.

Frouin, R., S. F. Iacobellis, and P.-Y. Deschamps, 2001: Influence of oceanic whitecaps on the global radiation budget. *Geophys. Res. Lett.*, **28**, 1523–1526, https://doi.org/10.1029/2000GL012657.

Gemmrich, J. R., M. L. Banner, and C. Garrett, 2008: Spectrally resolved energy dissipation rate and momentum flux of breaking waves. *J. Phys. Oceanogr.*, **38**, 1296–1312, https://doi.org/10.1175/2007JPO3762.1.

Gordon, H. R., and M. Wang, 1994: Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors. *Appl. Opt.*, **33**, 7754–7763, https://doi.org/10.1364/AO.33.007754.

Hanson, J. L., and O. M. Phillips, 1999: Wind sea growth and dissipation in the open ocean. *J. Phys. Oceanogr.*, **29**, 1633–1648, https://doi.org/10.1175/1520-0485(1999)029<1633:WSGADI>2.0.CO;2.

He, K., X. Zhang, S. Ren, and J. Sun, 2016: Identity mappings in deep residual networks. *Computer Vision–ECCV 2016*,

B. Leibe et al., Eds., Lecture Notes in Computer Science, Vol. 9908, Springer, 630–645.

Hooker, G., S. E. Brumer, C. J. Zappa, and E. C. Monahan, 2020: Inferences to be drawn from a consideration of power-law descriptions of multiple data sets each comprised of whitecap coverage, WB, and 10-m elevation wind speed measurements (U_{10}). *Recent Advances in the Study of Oceanic Whitecaps: Twixt Wind and Waves*, P. Vlahos and E. Monahan, Eds., Springer, 43–63.

Hwang, P. A., and M. A. Sletten, 2008: Energy dissipation of wind-generated waves and whitecap coverage. *J. Geophys. Res.*, **113**, C02012, https://doi.org/10.1029/2007JC004277.

Jadon, S., 2020: A survey of loss functions for semantic segmentation. *2020 IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Via del Mar, Chile, Institute of Electrical and Electronics Engineers, 1–7, https://doi.org/10.1109/CIBCB48159.2020.9277638.

Jia, N., and D. Zhao, 2019: The influence of wind speed and sea states on whitecap coverage. *J. Ocean Univ. China*, **18**, 282–292, https://doi.org/10.1007/s11802-019-3808-7.

Kleiss, J. M., and W. K. Melville, 2011: The analysis of sea surface imagery for whitecap kinematics. *J. Atmos. Oceanic Technol.*, **28**, 219–243, https://doi.org/10.1175/2010JTECHO744.1.

Koepke, P., 1984: Effective reflectance of oceanic whitecaps. *Appl. Opt.*, **23**, 1816, https://doi.org/10.1364/ao.23.001816.

Kraan, G., W. A. Oost, and P. A. E. M. Janssen, 1996: Wave energy dissipation by whitecaps. *J. Atmos. Oceanic Technol.*, **13**, 262–267, https://doi.org/10.1175/1520-0426(1996)013<0262:WEDBW>2.0.CO;2.

Liu, M., and B. Yang, 2022: Evaluation of sea surface temperature-dependent whitecap coverage parameterizations using in situ data. *Ocean Sci. J.*, **57**, 174–185, https://doi.org/10.1007/s12601-022-00060-4.

Long, J., E. Shelhamer, and T. Darrell, 2015: Fully convolutional networks for semantic segmentation. *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, Institute of Electrical and Electronics Engineers, 3431–3440, https://doi.org/10.1109/CVPR.2015.7298965.

Lyu, M., H. Potter, and C. O. Collins, 2021: The impacts of gustiness on air–sea momentum flux. *Fluids*, **6**, 336, https://doi.org/10.3390/fluids6100336.

Malila, M. P., J. Thomson, Ø. Breivik, A. Benetazzo, B. Scanlon, and B. Ward, 2022: On the groupiness and intermittency of oceanic whitecaps. *J. Geophys. Res. Oceans*, **127**, e2021JC017938, https://doi.org/10.1029/2021JC017938.

Marmorino, G. O., and G. B. Smith, 2005: Bright and dark ocean whitecaps observed in the infrared. *Geophys. Res. Lett.*, **32**, L11604, https://doi.org/10.1029/2005GL023176.

Minaee, S., Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, 2021: Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 3523–3542, https://doi.org/10.1109/TPAMI.2021.3059968.

Mironov, A. S., and V. A. Dulov, 2008: Detection of wave breaking using sea surface video records. *Meas. Sci. Technol.*, **19**, 015405, https://doi.org/10.1088/0957-0233/19/1/015405.

Monahan, E. C., 1971: Oceanic whitecaps. *J. Phys. Oceanogr.*, **1**, 139–144, https://doi.org/10.1175/1520-0485(1971)001<0139:OW>2.0.CO;2.

——, and I. G. O'Muircheartaigh, 1986: Whitecaps and the passive remote sensing of the ocean surface. *Int. J. Remote Sens.*, **7**, 627–642, https://doi.org/10.1080/01431168608954716.

——, and D. K. Woolf, 1989: Comments on "variations of whitecap coverage with wind stress and water temperature. *J. Phys. Oceanogr.*, **19**, 706–709, https://doi.org/10.1175/1520-0485(1989)019<0706:COOWCW>2.0.CO;2.

Padmanabhan, S., S. C. Reising, W. E. Asher, L. A. Rose, and P. W. Gaiser, 2006: Effects of foam on ocean surface microwave emission inferred from radiometric observations of reproducible breaking waves. *IEEE Trans. Geosci. Remote Sens.*, **44**, 569–583, https://doi.org/10.1109/TGRS.2006.870234.

Potter, H., G. B. Smith, C. M. Snow, D. J. Dowgiallo, J. P. Bobak, and M. D. Anguelova, 2015: Whitecap lifetime stages from infrared imagery with implications for microwave radiometric measurements of whitecap fraction. *J. Geophys. Res. Oceans*, **120**, 7521–7537, https://doi.org/10.1002/2015JC011276.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab et al., Eds., Lecture Notes in Computer Science, Vol. 9351, Springer International Publishing, 234–241.

Sáez, F. J., P. A. Catalán, and C. Valle, 2021: Wave-by-wave nearshore wave breaking identification using U-Net. *Coastal Eng.*, **170**, 104021, https://doi.org/10.1016/j.coastaleng.2021.104021.

Salisbury, D. J., M. D. Anguelova, and I. M. Brooks, 2013: On the variability of whitecap fraction using satellite-based observations. *J. Geophys. Res. Oceans*, **118**, 6201–6222, https://doi.org/10.1002/2013JC008797.

Scanlon, B., Ø. Breivik, J.-R. Bidlot, P. A. E. M. Janssen, A. H. Callaghan, and B. Ward, 2016: Modeling whitecap fraction with a wave model. *J. Phys. Oceanogr.*, **46**, 887–894, https://doi.org/10.1175/JPO-D-15-0158.1.

Schwendeman, M., and J. Thomson, 2015a: Observations of whitecap coverage and the relation to wind stress, wave slope, and turbulent dissipation. *J. Geophys. Res. Oceans*, **120**, 8346–8363, https://doi.org/10.1002/2015JC011196.

——, and ——, 2015b: A horizon-tracking method for shipboard video stabilization and rectification. *J. Atmos. Oceanic Technol.*, **32**, 164–176, https://doi.org/10.1175/JTECH-D-14-00047.1.

Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Siddique, N., P. Sidike, C. Elkin, and V. Devabhaktuni, 2020: U-Net and its variants for medical image segmentation: Theory and applications. arXiv, 2011.01118v1, https://doi.org/10.48550/arXiv.2011.01118.

Simonyan, K., and A. Zisserman, 2014: Very deep convolutional networks for large-scale image recognition. arXiv, 1409.1556v6, https://doi.org/10.48550/arXiv.1409.1556.

Stramska, M., and T. Petelski, 2003: Observations of oceanic whitecaps in the north polar waters of the Atlantic. *J. Geophys. Res.*, **108**, 3086, https://doi.org/10.1029/2002JC001321.

Sugihara, Y., H. Tsumori, T. Ohga, H. Yoshioka, and S. Serizawa, 2007: Variation of whitecap coverage with wave-field conditions. *J. Mar. Syst.*, **66**, 47–60, https://doi.org/10.1016/j.jmarsys.2006.01.014.

Veron, F., C. Hopkins, E. L. Harrison, and J. A. Mueller, 2012: Sea spray spume droplet production in high wind speeds. *Geophys. Res. Lett.*, **39**, L16602, https://doi.org/10.1029/2012GL052603.

Voulodimos, A., N. Doulamis, A. Doulamis, and E. Protopapadakis, 2018: Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.*, **2018**, 7068349, https://doi.org/10.1155/2018/7068349.

Wang, Y., Y. Sugihara, X. Zhao, H. Nakashima, and O. Eljamal, 2020: Deep learning-based image processing for whitecaps on

the ocean surface. *J. Japan Soc. Civ. Eng.*, **76B2**, I_163–I_168, https://doi.org/10.2208/kaigan.76.2_I_163.

Wu, J., 1988: Variations of whitecap coverage with wind stress and water temperature. *J. Phys. Oceanogr.*, **18**, 1448–1453, https://doi.org/10.1175/1520-0485(1988)018<1448:VOWCWW>2.0.CO;2.

Yang, X., and H. Potter, 2021: A novel method to discriminate active from residual whitecaps using particle image velocimetry. *Remote Sens.*, **13**, 4051, https://doi.org/10.3390/rs13204051.

Yeung, M., E. Sala, C.-B. Schönlieb, and L. Rundo, 2022: Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.*, **95**, 102026, https://doi.org/10.1016/j.compmedimag.2021.102026.

Zhou, S., F. Xu, and R. Shi, 2022: Whitecap fraction parameterization and understanding with deep neural network. *Remote Sens.*, **15**, 241, https://doi.org/10.3390/rs15010241.