Journal of Advanced Marine Engineering and Technology, Vol. 48, No. 5, pp. 346~352, 2024
J. Advanced Marine Engineering and Technology (JAMET)
https://doi.org/10.5916/jamet.2024.48.5.346

ISSN 2234-7925 (Print)
ISSN 2765-4796 (Online)
Original Paper

Whitecap detection analysis using instance segmentation models

Min-Seo Kim[1] · Ju-Hyeon Seong[†]

(Received September 21, 2024 ; Revised October 5, 2024 ; Accepted October 18, 2024)

Abstract: In the ocean, whitecaps resulting from various factors such as wind and ships induce considerable disturbances in image-based object detection and recognition models. These disturbances limit the accuracy of these models. Previous studies have employed object detection models to detect and remove irregular phenomena of nonuniform sizes and shapes. Because these models are trained on stereotyped rectangular objects, expressing the exact shape of an actual object remains challenging. Therefore, to ensure the accuracy of real-time optimal whitecap detection in a marine environment, we applied and analyzed core instance segmentation models based on real images captured by a drone. The instance segmentation models used in the experiment were selected as two models, namely anchor- and anchor-free models, and four models were analyzed to ensure real-time accuracy and processing speed.

Keywords: Whitecaps, Instance Segmentation model, Anchor-based model, Anchor-free based model, Drone video

1. Introduction

Advancement in artificial intelligence (AI) technology has resulted in technological convergence toward AI in almost all fields. In computer vision, this convergence is being adopted into various environments based on the environmental robustness of machine learning. Artificial intelligence (AI) is increasingly being incorporated in the marine industry. However, directly applying terrestrial technologies in marine environments is challenging because of the difficulty of data collection and weather complexity **[1]**. Whitecaps **[2][3]** can occur when strong winds cause the crests of waves to break or when persistent winds in one direction result in strong waves. Therefore, whitecaps can be indicators for wind strength and direction. Furthermore, whitecaps can occur when waves hit a reef, indicating shallow water or the presence of underwater obstacles, providing critical navigation information. Furthermore, observing whitecaps caused by schools of fish can be used for fishing operations. Thus, whitecap detection provides insights into real-time wind and sea conditions, underwater topography, and fish movement and highlight the necessity for continuous observations. Whitecaps observed through video are surface-level phenomena that occur in oceans. When combined with systems such as radar, which can analyze the environment below the surface, this phenomenon enables comprehensive analysis.

Because whitecaps are caused by numerous natural interactions, their size, shape, and ratio are not constant. Because of the irregularity of whitecaps, existing object detection models detect objects by setting areas of fixed sizes and ratios, rendering learning the characteristics of objects accurately difficult. Furthermore, the color of the sea, which changes depending on the weather and climate during classification tasks, is a major limitation.

Despite these limitations, numerous methods have been proposed for detecting events and objects, such as coastal boundaries and port surveillance, occurring at sea. Studies **[4][5][6]** have improved the object detection accuracy at sea by combining general object detection with white-wave detection. Hu *et al.* **[4]** applied image post-processing techniques and anomalous detection to address the problem that the amount of sunlight considerably affects white-wave detection performance. Although this method can perform precise white-wave detection, the method has a complex structure for detecting outliers and requires considerable computation, rendering real-time detection difficult. Atkin *et al.* **[5]** applied YOLOv5, an object-detection model, to analyze the quality of surfing waves. However, this method has a low detection resolution for irregular white waves, rendering multi-detection difficult in environments in which many white waves occur. Vrecica *et al.* **[6]** proposed a method for detecting

[†] Corresponding Author (ORCID: http://orcid.org/0000-0002-8198-0439): Professor, Division of Maritime AI and Cyber Security & Interdisciplinary Major of Maritime AI Convergence, Korea Maritime & Ocean University, 727, Taejong-ro, Yeongdo-gu, Busan 49112, Korea, E-mail: jhseong@kmou.ac.kr, Tel: 051-410-5031

[1] M. S., Department of Maritime AI and Cyber security & Interdisciplinary Major of Maritime AI Convergence, National Korea Maritime & Ocean University, E-mail: sea7616@g.kmou.ac.kr, Tel: 051-410-7624

whitecaps in images using U-Net, which is a segmentation model. A deep-learning-based segmentation model was applied to detect whitecaps. However, real-time processing is complex because of U-Net, which exhibits processing speed problems. Processing speed and accuracy problems could be attributed to limitations in improving the recognition accuracy caused by the structural method of the object detector and the complexity of the segmentation model being applied. Therefore, to commercialize maritime irregular whitecap detection technology, lightweight segmentation technology should be applied to guarantee real-time performance.

In this study, based on actual drone images, we analyzed various state-of-the-art instance-segmentation-based white-wave detection models to ensure optimal white-wave detection accuracy and real-time performance in a maritime detection environment.

We analyzed the results by applying two anchor-based and two anchor-free methods, depending on the presence or absence of anchors in the instance segmentation model.

2. Related Works

2.1 Anchor-Based Segmentation Models

Segmentation techniques include semantic segmentation **[7][8]**, which distinguishes the classes of all objects in an image, and instance segmentation **[9][10][11][12]**, which distinguishes only specific objects of classes in an image. Early segmentation directly estimates the region of an object in an image. However, new models have applied an object detector. In contrast to detectors that estimate the location of an object using a bounding box, instance segmentation outputs a region in pixel units to identify the shape of an object. Similar to object detection, segmentation has two approaches, namely anchor-based **[9][10]** and anchor-free methods **[11][12]**, depending on the presence or absence of an anchor. The anchor, which is the method proposed for object detection, can be used for bounding box, and the model achieves high learning efficiency by setting this anchor in advance. Mask R-CNN, a representative method of instance segmentation, is a two-stage segmentation model that is designed to enable faster segmentation by applying a parallel mask branch for mask prediction to the branch of the regressor based on the object detection model. Because these two stages are classified into object detection and semantic segmentation, each stage incorporates an independent method. This model exhibits high segmentation performance by replacing the RoI pooling method with RoIAlign to predict accurate pixel masks and resolve misalignments between features and RoIs.

TensorMask **[10]**, a representative one-stage segmentation model that combines object detection and semantic segmentation, improves the computational speed and segmentation performance of overlapping instances by using a four-dimensional (4D) tensor for dense instance detection.

Anchor-based segmentation models can obtain stable and highly accurate results by segmenting candidate regions. However, these models are highly dependent on the anchors generated in advance, which renders the detection of significant changes in the size or ratio of an object difficult.

2.2 Anchor Free-Based Segmentation Models

Existing anchor-based segmentation models exhibit insufficient usability because of fixed-size anchors and additional parameter tuning. Therefore, anchor-free segmentation incorporates Fully Convolutional One-stage Object detection (FCOS) **[13]** instead of anchors that pregenerate object candidate regions. FCOS extracts feature maps of various sizes by using the Feature Pyramid Network (FPN) **[14]** structure. Each feature map finds the object's center point to be detected and limits the overgeneration of candidate object regions. Because this model sets the region based on the center point, the region display standard is not the box starting point but the center point and the lengths of the left, right, top, and bottom from the center point.

CenterMask, a representative one-stage segmentation model incorporating FCOS, improves segmentation performance through a Spatial Attention-Guided Mask (SAG-Mask) and Spatial Attention Module (SAM).

Representative instance segmentation methods, such as Mask R-CNN, require ROI operations. The axis-aligned features of the existing RoIs can result in excessive computational consumption when objects in the input image exhibit irregular shapes. Conditional convolutions for instance segmentation (CondInst) **[12]**, an anchor-free-based segmentation model, solves instance segmentation with an FCN [7]. Instead of using an ROI, in this model, a dynamic instance recognition network conditioned on instances is used such that the mask head is compact and the processing speed is fast.

3. Experimental Environment Configuration Based on Whitecaps Analysis

3.1 Characteristics Analysis of Whitecaps

We conducted a comprehensive analysis based on the most commonly used square-shaped patches for AI detection to determine whether we could distinguish oceans and whitecaps.

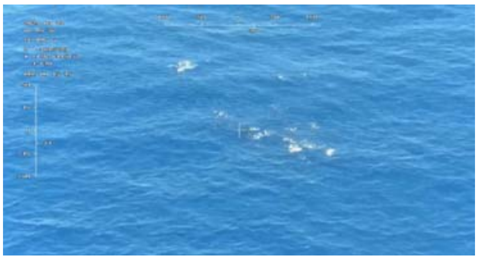


Figure 1: Whitecaps video image captured by a drone

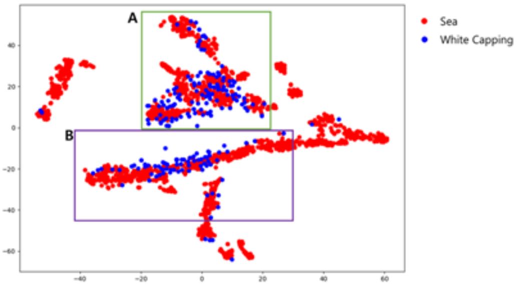


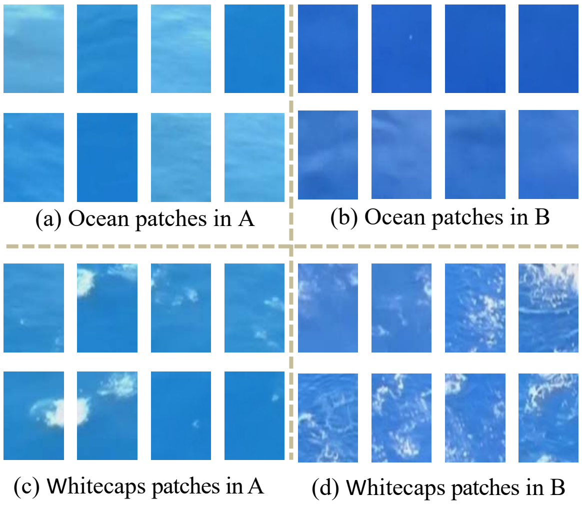
Figure 2: Results of applying DBSCAN using patched images of the sea and whitecaps

Figure 1 depicts a whitecap video captured by a drone. Whitecaps, a natural phenomenon, exhibit irregular characteristics in terms of their appearance, size, and other characteristics. Furthermore, depending on weather conditions, such as light reflection, illuminance, and brightness, distinguishing the background ocean and whitecaps can be difficult. Therefore, we applied DBSCAN **[15]** to analyze whether white waves could be distinguished from the ocean as a single data feature.

Figure 2 depicts the results of applying DBSCAN to images of the ocean and whitecaps. The images were captured using a drone at 30 frame per second (FPS) for 15 s, and the camera UI and sky region were removed through preprocessing. To divide the images into small sections of the ocean and whitecaps, we divided them into 37,800 patch images using patches of 100 × 100 pixels and used them for analysis.

Red dots represent ocean patches and blue dots represent patches containing whitecaps. First, the distribution of ocean patches was observed as a result of multiple clusters, and the clusters were also low in density, rendering distinguishing ocean features difficult.

The results of analyses have attributed this phenomenon to the color values varying depending on the shooting time and location due to sea waves, differences in illuminance, among other factors. Therefore, each ocean patch shares a single characteristic. The whitecap patch (blue dot) shares some similarities with the ocean patch. Thus, distinguishing it as a characteristic unique to whitecaps is challenging. Thus, the images were visually analyzed by dividing them into regions A and B, where whitecap data distribution appeared.



(a) Ocean patches in A (b) Ocean patches in B

(c) Whitecaps patches in A (d) Whitecaps patches in B

Figure 3: Example of images clustered into regions A and B: (a) the ocean patch in A, (b) whitecaps patch

Figure 3 depicts an example of the images clustered into regions A and B. (a) and (c) represent some of the ocean patches and whitecaps in cluster A, respectively. (b) and (d) represent ocean patches and whitecaps in the Cluster B region, respectively.

Among images in (a) and (c), some images are perfectly blue; however, irregular afterimages are caused by light reflection and waves. Compared with images in (b) and (d), these afterimages can be recognized as different parts with different colors. Furthermore, they have the same characteristics because of their irregularity. However, a white color patch is distinctive.

By distinguishing and cluster-white waves and the sea using DBSCAN, whitecap patches can be clustered to some extent by features such as color. However, distinguishing these from white waves when sea patches include white waves and light reflections is difficult. Thus, applying techniques such as the CNN **[16]** and object detection **[17]**, which necessarily include background learning, is difficult.

This phenomenon distorts detection accuracy. Because the appearance of whitecaps always changes, standardization is difficult. However, if the background blue color is removed and only the focus is on the outer shape of the white waves, detection accuracy can be improved.

These results determined that a segmentation model that could learn the irregular shape information of whitecaps would be suitable for white-wave detection, and various models were applied.

3.2 Experimental Environment

To comparatively analyze the performances of the segmentation

Table 1: Composition of the dataset used for segmentation-based whitecaps

Item	Images
Train	714
Test	92
Validation	214

models, data acquisition and experimental environments were configured as follows:

Table 1 lists the composition of the dataset used for segmentation-based whitecaps. We divided the drone video into 714 learning images, 92 test images, and 214 validation images for use as a learning and performance detection dataset. A dataset was constructed using the correct answer data for each image.

To acquire the data, the drone used 0000 and filmed near the National Korea Maritime and Ocean University drone airfield, which is a nationally designated drone airspace. A GPU (NVIDIA A100 80 GB PCle × 1) was used for AI training and testing of the whitecap detection model.

3.3 Selection of Experimental Segmentation Models

We selected four instance segmentation models, rather than detection models, through experiments on the characteristics of whitecaps. **Figure 4** presents a systematic table of instance segmentation models. We conducted experiments by selecting representative models for each category of instance segmentation models to evaluate the whitecap detection performance of the instance segmentation models. However, the bottom-up method was excluded from these experiments. Because the bottom-up method works well when regular features are clear, applying the method to detect whitecaps, which are irregular objects, is difficult.

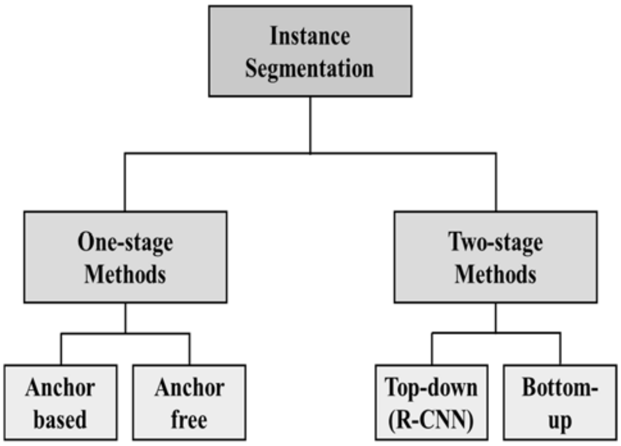


Figure 4: Systematic table of instance segmentation models

Mask R-CNN, a representative instance segmentation model of the R-CNN series, exhibits high detection performance and fast detection speed. This model can perform detailed segmentation of candidate regions with a Region Proposal Network (RPN) by using an anchor. Furthermore, the model can be used to perform detailed object detection because the model can perform multiple tasks, including object detection, bounding box regression, and segmentation mask prediction simultaneously. Therefore, the model is suitable for small-sized whitecap detection because of its high detection performance for small objects. Furthermore, the Region of Interest alignment (RoIAlign) proposed in the Mask R-CNN model allows for a precise extraction of object boundaries and enhances adaptability to changes in the object size. These advantages can be used for whitecap detection because the model is well suited for detecting complex whitecap regions. Furthermore, because of the wide coverage of drone cameras, objects of varying sizes are captured depending on the distance. The ability of the model to handle such scale variability renders the model suitable for experimental models.

TensorMask, the proposed model for performing dense segmentation, represents masks as high-dimensional tensors that can effectively segment objects such as whitecaps, which are difficult to detect because of overlapping objects, unclear outlines, or unclear shapes. Furthermore, Tensormask's dense sliding-window approach can be used to predict masks across the entire image area, rendering the model suitable for whitecap detection through light features mixed with the background over a large area.

CenterMask is a segmentation model based on the FCOS detector. Because FCOS is a pixel-wise prediction of objects, FCOS can effectively separate complex backgrounds from objects.

Therefore, when the background is unstable, as in the case of the ocean, it can be effective in separating the background from objects. Furthermore, the model introduces SAG-Mask for segmenting the candidate regions extracted by the detector. In the segmentation process, spatial attention was introduced to emphasize the characteristics of the objects in the region. Therefore, this spatial attention method is suitable for cases in which separating the background and object, such as whitecaps, is difficult.

Finally, unlike existing instance-segmentation models, CondInst is an FCN structure that does not generate an RoI. which is the detection result of the segmentation model. Because whitecaps have large irregularities in size and ratio, whitecap detection may not be possible if they do not fit the size and ratio of an anchor box of a prespecified size. Therefore, the structure of CondInst, which does not generate an RoI, is suitable for whitecap detection.

4. Experimental results

4.1 Experiment Result

Table 2: Classification criteria by the object size in COCO

Item	Scales (Pixels)
Small	area < 32^2
Medium	32^2 < area < 96^2
Large	area > 96^2

We analyzed a suitable white-wave detection segmentation model based on the characteristics and results of the previously selected models. The whitecap detection performance of the model was evaluated by dividing it into the mean average precision (mAP), AP50, and AP75. **Table 2** lists the criteria for classifying object sizes in COCO. Because the evaluation criteria for the white waves of various sizes differ depending on their sizes, the performance was analyzed by classifying them into small, medium, and large objects by applying the classification criteria of COCO.

Tables 3 and **4** present the box and segmentation AP results for each model, respectively. In the case of the Mask R-CNN model, unlike other models that exhibit differences in detection performance by item in terms of detection accuracy by the object size,

AP_s, AP_m, and AP_l consistently exhibited high performance. The small whitecaps exhibit clear characteristics in small areas, and with the increase in the size, the whitecap boundary is not clear, resulting in considerable interference from the background. Therefore, Mask R-CNN and CenterMask2, which consistently recorded high performances by size, were models that could evenly respond to both characteristics. For large whitecaps, boundaries are complex and ambiguous.

Therefore, the CondInst model, including CenterMask2, can effectively detect objects with such characteristics. FCOS, an anchor-free detection model, was used to investigate the cause for this phenomenon. Unlike anchor methods such as Mask R-CNN, the FCOS

model does not create an area in the form of a box, but determines the center point and subsequently estimates the area of the object from the center point. Therefore, this method is effective for detecting objects such as large white waves, where accurately defining the object range is difficult.

Figure 5 depicts an example of the application of the results of the four experimental models. (a) Visualization of the learning results of the Mask R-NN model. As presented in **Tables 3** and **4**, white wave detection performed using photographs achieved detection performance, including small object detection. (b) is an image showing the application results of the TensorMask model. In the case of the TensorMask model, the detection and segmentation performance for small objects was poor, and the results were reflected in the visualization photographs. In this model, detection of small white waves is difficult. By contrast, large object detection achieved high detection performance, and segmentation from the background was performed well.

(c) Visualization of the application results of the Centermask2 model. Centermask2 revealed the highest overall detection performance, with Mask R-CNN. The visualization results also revealed that detection was performed evenly from small to large white waves.

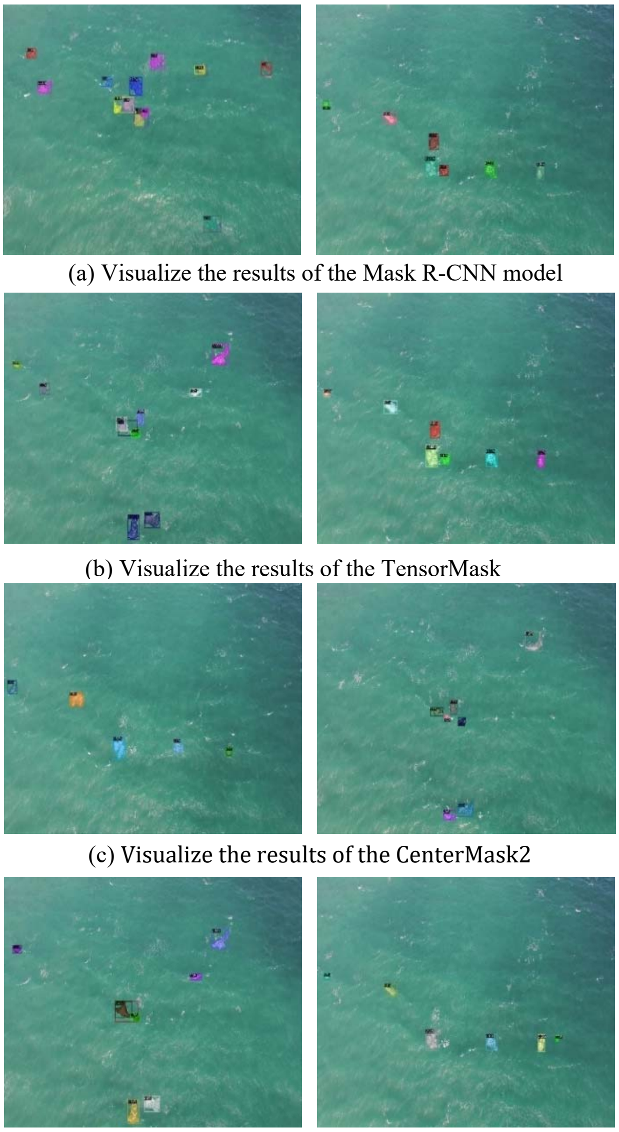
Here, (d) is a visualization of the application result of the CondInst Similar to the TensorMask model, the CondInst model did not perform well in detecting small objects. However, the model could detect large white waves accurately.

Table 3: Instance segmentation and detection performance: Bounding box AP

Models	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Mask R-CNN	46.9	77.1	49.3	46.5	46.1	58.1
Tensor Mask	44.9	78.4	45.9	37.3	45.5	60.4
CenterMask	45.7	75.1	48.7	43.0	46.4	62.5
CondInst	46.3	73.7	50.0	42.4	47.3	61.7

Table 4: Instance segmentation and detection performance: Segmentation AP

Models	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Mask R-CNN	41.0	77.2	36.8	25.7	42.6	59.5
Tensor Mask	30.9	71.4	19.4	9.4	34.8	62.2
CenterMask	42.0	79.7	38.5	23.8	42.9	60.3
CondInst	36.6	76.1	29.0	13.5	38.8	59.7



(a) Visualize the results of the Mask R-CNN model

(b) Visualize the results of the TensorMask

(c) Visualize the results of the CenterMask2

(d) Visualize the results of the CondInst

Figure 5: Visualization images resulting from result of the segmentation model

In terms of overall performance indicators, Mask R-CNN revealed the highest object detection performance, and CenterMask2 revealed the highest segmentation performance. Therefore, we concluded that the Mask R-CNN model is preferred when accurate object detection performance is a priority, and the CenterMask2 model is preferred when region segmentation is a priority.

Table 5 lists the detection speeds of the model. The inference time is the time required to extract results for one image, and the FPS is the number of images that can be processed per second. This value is calculated based on the average time of the entire inference time. The CondInst model achieved the fastest detection speed in the experiment, whereas the Mask R-CNN model achieved the slowest detection speed.

Table 5: FPS and inference time per model

Models	FPS	Inference time(ms)
Mask R-CNN	8.33	0.12
Tensor Mask	9.62	0.14
CenterMask	9.83	0.12
CondInst	10.71	0.11

5. Conclusion

In this study, we analyzed whitecap data and developed a suitable segmentation model to detect the white-wave phenomenon through instance segmentation. The experimental results revealed that whitecap detection using the segmentation model performed well, but detection was difficult in some situations, such as when the white wave size was very small or when the whitecaps were scattered and features were faint. In a follow-up study, we will conduct additional research to identify the characteristics of clear separation between the background and whitecaps by referring to the highly variable color of the sea through density-based clustering.

The dataset used in this study was limited to specific conditions, such as time, location, and causes of whitecaps at the time of capture. Therefore, to address the highly variable ocean environment that fluctuates with weather and location, in the future, the dataset should be expanded to include whitecap data captured under various conditions. By obtaining an expanded dataset categorized by the causes of whitecaps, detailed analysis based on the cause of whitecaps should be conducted.

Acknowledgement

This paper was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (RS-2024-00424595), and National IT Industry Promotion Agency(NIPA) grant funded by the Korea Government(MOTIE) (S2001-24-1006).

Author Contributions

Conceptualization, M. S. Kim and J. -H. Seong; Methodology, M. S. Kim; Software, M. S. Kim; Validation, M. S. Kim and J. -H. Seong; Formal Analysis, M. S. Kim; Investigation, M. S. Kim; Resources, M. S. Kim; Data Curation, M. S. Kim; Writing—Original Draft Preparation, M. S. Kim; Writing—Review & Editing, J. -H. Seong; Visualization, M. S. Kim; Supervision, J. -H. Seong; Project Administration, J. -H. Seong; Funding Acquisition, J. -H. Seong.

References

[1] C. E. Stringari, et al., “Deep neural networks for active wave breaking classification,” Scientific Reports, vol. 11, no. 1, 3604, 2021.

[2] Y. Sugihara, H. Tsumori, T. Ohga, H. Yoshioka, and S. Serizawa, “Variation of whitecap coverage with wave-field conditions,” Journal of Marine Systems, vol. 66, no. 1-4, pp. 47-60, 2007.

[3] B. Scanlon and B. Ward, “Oceanic wave breaking coverage separation techniques for active and maturing whitecaps,” Methods in Oceanography, vol. 8, pp.1-12, 2013.

[4] X. Hu, Q. Yu, A. Meng, C. He, S. Chi, M. Li, “Using optical flow trajectories to detect whitecaps in light-polluted videos,” Remote Sensing, vol. 14, no. 22, 5691, 2022.

[5] E. A. Atkin, “Machine-learned peel angles for surfing wave quality monitoring,” Australasian Coasts & Ports 2021: Te Oranga Takutai, Adapt and Thrive: Te Oranga Takutai, Adapt and Thrive, Christchurch, NZ: New Zealand Coastal Society, pp. 91-97, 2022.

[6] T. Vrecica, Q. Paletta, and L. Lenain, “Deep learning applied to sea surface semantic segmentation: Filtering sunglint from aerial imagery,” Proceedings of the ICML 2021 Workshop on Tackling Climate Change with Machine Learning, vol. 23, 2020.

[7] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, 2015.

[8] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6230-6239, 2017.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.

[10] X. Chen, R. Girshick, K. He, and P. Dollar, “Tensormask: A foundation for dense object segmentation,” Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2061-2069, 2019.

[11] Y. Lee and J. Park, “Centermask: Real-time anchor-free instance segmentation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp: 13906-13915, 2020.

[12] Z. Tian, C. Shen, and H. Chen, “Conditional convolutions for instance segmentation,” Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer International Publishing, pp. 282-298, 2020.

[13] Z. Tian, C. Shen, H. Chen, and T. He, “FCOS: A simple and strong anchor-free object detector,” IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 44, no. 4, pp. 1922-1933, 2022.

[14] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 936-944, 2017.

[15] K. Khan, et al., “DBSCAN: Past, present and future,” The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014) IEEE, pp. 232-238, 2014.

[16] M. Shafiq and Z. Gu, “Deep residual learning for image recognition: A survey,” Applied Sciences, vol. 12, no. 18, 8972, 2022.

[17] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” Proceedings of the IEEE vol. 111, no. 3, pp. 257-276, 2023.

[18] W. Gu, S. Bai, and L. Kong, “A review on 2D instance segmentation based on deep neural networks,” Image and Vision Computing, vol. 120, 104401, 2022.

[19] T. Lin, *et al.*, “Microsoft coco: Common objects in context,” Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.