Segmentation of seagrass (*Enhalus acoroides*) using deep learning mask R-CNN algorithm

NA Lestari, I Jaya* and M Iqbal
Department of Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University, Dramaga, Bogor 16680, West Java, Indonesia

*E-mail: indrajaya@apps.ipb.ac.id

Abstract. Seagrass is an Angiosperms that live in shallow marine waters and estuaries. The method commonly used to identify seagrass is Seagrass-Watch which is done by sampling seagrass or by carrying a seagrass identification book. Technological developments in the era of the industrial revolution 4.0 made it possible to identify seagrass automatically. This research aims to apply the deep learning algorithm to detect seagrass recorded by underwater cameras. *Enhalus acoroides* seagrass species identification was carried out using a deep learning method with the mask region convolutional neural networks (Mask R-CNN) algorithm. The steps in the research procedure include collecting, labeling, training, testing models, and calculating the seagrass area. This study used 6000 epochs and got a measure of value generated by the model of ± 1.2. The Precision value, namely the model's ability to correctly classify objects, reached 98.19% and the model's ability to find all positive objects, based on system testing was able to perform recall is 95.04% and the F1 Score value of 96.58%. The results showed that the MASK R-CNN algorithm could detect and segment seagrass *Enhalus acoroides*.

Keywords: deep learning, *Enhalus acoroides*, Mask R-CNN, seagrass

1. Introduction
Seagrass is an Angiospermae that lives in shallow marine waters and estuaries. Seagrass plants consist of leaves and sheaths, creeping stems which are usually called rhizomes and roots that grow on the rhizome [1]. A shallow stretch of sea dominated by seagrass is known as a seagrass meadow. Seagrass beds can consist of single or mixed types of seagrass vegetation. Seagrass beds are places for various types of fish to shelter, find food, lay eggs, and raise their young [2]. The functions and benefits of seagrass beds in shallow water ecosystems are primary producers, bottom stabilizers, sediment catchers also nutrient recyclers [3]. From about 60 species of seagrass known in the world, Indonesia has about 13 species [4]. One type of seagrass that dominates Indonesian waters is *Enhalus acoroides*.

Seagrass monitoring activities play an important role in managing the coastal environment for two reasons. This activity is a method for improving management practices and can provide information on the status and condition of seagrass beds [1]. A general description of the existence and knowing of seagrass species in an area can be done by identifying seagrasses. The method commonly used to identify seagrass is the Seagrass-Watch. The method was carried out using a 50m line transect from land to sea with 3 replications. The distance between transects is 50 meters, and put a square (50x50 cm$^2$) on the right side of the line with a space between the squares of 5 meters [5]. In addition, there is The Indonesian Institute of Sciences (LIPI) method, which is a quadratic transect (perpendicular to the
shoreline) that is modified from the Seagrass-Watch method [1]. Identification of seagrass using that method is made by taking seagrass samples or by bringing a seagrass identification book so that it takes a relatively long time to identify.

Technological developments in the industrial revolution 4.0 era made it possible to identify seagrasses automatically. The method developed is to use artificial intelligence, where computers can automatically recognize visible objects. The computer is implanted with a program to carry out learning to recognize the object to be identified. Learning on the computer can use the deep learning model [6]. Deep learning is defined as an artificial neural network that has many layers to imitate the workings of the human brain [7, 8]. Deep learning significantly improves the “state of the art” in speech recognition, visual object recognition, object detection, and many other domains [8, 9]. This method is proven to be effective in identifying patterns from data or objects. This is because this method can predict existing objects and their location [10, 11].

This research uses seagrass, which results from underwater camera recording as an object to be detected by the Mask R-CNN algorithm. Mask Region Convolutional Neural Networks (Mask RCNN) framework has a simple and flexible concept for efficiently detecting and segmenting objects. The method of Mask R-CNN is a development of the previous process, namely Faster R-CNN. Faster R-CNN has two outputs for each object, namely class label and bounding box. In Mask, R-CNN added a third output, namely the mask object [12]. The R-CNN mask was chosen because it is not only capable of detecting but also segmenting. The segmentation performed by Mask R-CNN is instance segmentation. Instance segmentation is a process of separating a complete image into several pixel groups, which can then be categorized. Instance segmentation can localize different classes of object instances in various images [13]. The use of instance segmentation can distinguish each seagrass in the recorded image so that individual calculations and individual seagrass cover areas can be carried out.

2. Materials and methods
The proposed approach for *Enhalus acoroides* detection and segmentation is performed in five steps: data collection, data labeling, training dataset, validation, and calculation of seagrass coverage area.

2.1. Data collection
The dataset used in this research are pictures of the seagrass *Enhalus acoroides* obtained through underwater camera recordings from Mahesa Glagah Agung Satria on Beralas Pasir, Bintan [14]. The recording is 19 minutes long with a resolution of 640x480 and a framerate of 30 fps which is then extracted to obtain an image for each frame. The extraction process produces 20.823 images and continues with the data selection process. Data selection is carried out to get the best quality image. The image is not blurry so the seagrass at image is visible. In the data selection process, 5000 images were obtained which will be used for learning Mask R-CNN algorithm. The images used are divided by a ratio of 80:10:10. Where 400 for the train set and 50 for the test set so that a total of 450. While the other 50 images are used for model evaluation.

2.2. Data labeling
The following procedure after dataset collection is the data labeling process. The purpose of image labeling is to provide a label on the image according to the seagrass to be recognized in the training process. This process is carried out using labelme software. The labeling process is done by digitizing the seagrass image using the polygon tool and then labeling it according to the identified object. Label name used for the training process is *Enhalus acoroides*. The data that have been tagged are then stored in javascript object notation (JSON). An example of image labeling can be seen in Figure 1.

2.3. Training dataset
Dataset training is carried out to train the algorithm to recognize the dataset and form a model based on the training. Training is carried out through two processes, namely forward propagation and backward propagation. Forward propagation is a value processing from the input layer and produces the final value
at the output layer. At the same time, backpropagation is a backward propagation process to the layers behind it to improve the weight value generated in the forward process [6]. This process is carried out until the output is in the form of a file containing the weight and bias values. The process will be carried out continuously until it gets the production with the smallest loss value. This research uses Mask R-CNN technique. This technique is an algorithm used to detect objects, localize objects, and segment objects from an image [12]. The Mask R-CNN follows the Faster R-CNN model of the feature extractor followed by an operation known as ROI-Pooling to produce a standard-sized output suitable for input to the classifier, with three important modifications [15].

**Figure 2. Architecture Mask R-CNN.**

Figure 2 is an object detection architecture using Mask-RCNN which in the process, namely the input image, will enter the convolution process. In this process, feature extraction or removal of features from the image and a feature map are produced. The results obtained will then enter the fully connected layers and fully convolutional network layers. The result of a fully connected layer is a bounding box for object localization and probability or confidence class. While a fully convolutional network will produce segmentation so that the Mask R-CNN Mask is not only able to detect objects but also object
segmentation. The training process used by Backbone Resnet101 is because the accuracy of Resnet101 is better than other models. The batch used for optimization in the training process is 128. The batch size will affect the training process, and a large batch value will slow down the training process. The training process is carried out using 450 images with 6000 epochs and a 0.01 learning rate. The number of epochs is used to evaluate the model to recognize objects. Learning rate affects the intensity of the training process, effectiveness, and speed of the training process. The greater the value of the learning rate will make the training process faster. Still, it can exceed the optimal state or when it reaches the minimum error value so that the learning rate affects the network accuracy of a system [16].

2.4. Validation
The training test process is carried out using weights or bias weight values obtained from the training results that have been carried out. The input data used is an image of *Enhalus acoroides* seagrass, then the output produced in this process is a color segmentation on the seagrass that has been identified along with the name and presentation of the accuracy of the seagrass detection. The confusion matrix is used to measure the performance of the Mask R-CNN model which has previously been trained. The confusion matrix is a matrix that displays the predictions of the actual classification and classification [17]. The results obtained in the validation process can then be used to calculate the precision, recall, and *F1-Score* values of the model used. The values of precision, recall, and *F1-Score* are calculated using the equations shown below.

\[
Precision = \frac{TP}{TP + FP} \quad (1)
\]

\[
Recall = \frac{TP}{TP + FN} \quad (2)
\]

\[
F1-Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)
\]

The *TP* value is positive observations or seagrasses that are identified as correct. In contrast, *FP* has incorrectly identified seagrass or the number of positive data but is incorrectly classified by the system. *FN* is the number of negative data, namely seagrasses that are not specified, the accuracy of the system is calculated on the precision and recall values [18]. The precision value shows how precise the model is in determining the positive objects by following the actual. The recall value measures the model's ability to find all positive objects. The *F1-Score* is the harmonic average of *Precision* and *Recall* [19].

2.5 Calculation of seagrass coverage area
The seagrass coverage area is calculated to get the pixel value in the test data image before segmentation, which is then compared with the results after the segmentation process. This process is carried out using Adobe Photoshop software using pen tools to digitize the seagrass then the pixel value can be seen in the Photoshop histogram. In this case, the percentage of the area is calculated by comparing the number of seagrass pixels to the total number of seagrass pixels. In this case, the percentage of coverage area is calculated by equation. The seagrass coverage area calculation process can be seen in Figure 3.

\[
Percentage\ covering = \frac{Number\ of\ seagrass\ pixels}{Total\ pixels} \times 100\% \quad (4)
\]

3. Results and discussion

3.1. Dataset
The number of datasets used for training data is 450 images of seagrass *Enhalus acoroides* from underwater camera recording. The dataset is then given a label for each seagrass in the image. The seagrass images used have dimensions of 640 x 480, so that the number of pixels in each image is 307200 pixels. The total number of images used for the training and validation processes is 450 images.
Figure 3. Seagrass coverage area calculation process.

However, there can be more than one so that in the labeling process, each image can have more than one label. *Enhalus acoroides* can be up to 5 labels. The labeling criteria were all parts of the seagrass contained in the image with the labeling results obtained as many as 2,370 *Enhalus acoroides* seagrass labels. The more datasets collected, the better the object recognition process to be detected will be. This is because the model made will recognize the image to be identified more [20]. The dataset used must be varied and have good image quality to get training results with reasonable accuracy. Some pictures of the recorded seagrass can be seen in Figure 4.

Figure 4. Seagrasses from underwater camera recording.
3.2. Training loss

The training process carried out will produce a weight value that is used to detect seagrass. When the training process is carried out, the change in the loss value for each epoch can be seen. Epoch is one of the parameters used in the training process. Epoch is when the entire dataset has gone through the artificial neural network (ANN) training process until it is returned to the beginning in one round [21]. An epoch represents one cycle of a machine learning algorithm learning from the entire training data set. One epoch used for the training process means the machine learning algorithm has ‘learn’ from the training data as a whole. The iterative learning process aims to achieve convergence of the weighted values and depends on the number of epochs used [22]. The determination of the epoch value for the training process cannot be known with certainty, and it is necessary to do several experiments to get the epoch value with optimum results. The epoch value used will affect the training process when the epoch value used is greater, the time required for the training process will be longer than when using a smaller epoch value. The training results will be better if the loss value is getting smaller. The training loss graph can be seen in Figure 5.

![Training loss graph](image)

**Figure 5.** Training loss.

The loss value for this model starts with 2.4 and continues to fluctuate up and down until it reaches a loss value of 0.9 in the last epoch. How well the model detects and predicts the target can be known through training loss. Loss is a number that shows how bad the model prediction is, if the loss value is close to zero, it indicates that the model created has been able to recognize the object to be identified [6]. The smaller the loss value, the better the training results obtained [23]. The training loss graph obtained indicates that the model has been able to recognize the object to be identified.

3.3. Validation

The validation test of the training results is carried out using the weight value generated from the training process. Three results obtained in the validation process, including correctly identified seagrass, incorrectly identified seagrass, and unidentified seagrass, some examples of training results, can be seen in Figure 6 and Figure 7.

Figure 6 and Figure 7 show some of the correct identification results, unidentified seagrasses, and incorrectly identified seagrasses with the model that has been made. Seagrass identified correctly can be seen with the bounding box and labels and masks contained in the seagrass image. The absence can see seagrasses that are not identified of bounding boxes and labels and masks that mark the seagrass in the image. Seagrass identified incorrectly can be proven by bounding boxes and labels on the bounding box and masks that do not match the presence of seagrass in the image as in Figure 6, which identifies...
something but there is no seagrass in that part of the image. The results obtained indicate that the weight value obtained from the training dataset process can identify the seagrass that you want to identify. Seagrasses that can be identified but the identification results obtained are incorrect or the object is not seagrass incorrectly identified seagrass. Seagrasses that do not have a bounding box as a sign of seagrass to be identified are unidentified seagrass. This indicates that the model does not recognize the seagrass in the identified image. The validation results are carried out to automatically determine the results of the model that has been made for the identification of seagrasses. The results of the dataset test can be seen in Table 1.

![Figure 6. Seagrass identified correctly and seagrass unidentified.](image1)

![Figure 7. Seagrass identified incorrectly.](image2)

Table 1. Confusion matrix.

| Predicted label | Enhalus acoroides | Background |
|-----------------|-------------------|------------|
| Actual          |                   |            |
| Enhalus acoroides| 326               | 17         |
| Background      | 6                 | -          |

From Table 1, can see that 326 seagrasses were correctly identified, with 6 incorrectly and 17 were not identified. The higher the number of correctly identified seagrasses, the model could recognize the seagrasses to be identified. This error can be caused by the lack of datasets at certain positions, making it difficult for the model to identify seagrasses. Errors from the results obtained, both identification and unidentified errors can be overcome by collecting more varied datasets. The more data used for the training process means that more seagrass data will be introduced to the model. The more models recognize various forms of seagrass that will be identified, the better the results obtained from the identification with this deep learning model [6]. The results in Table 2. are used to test the model where the seagrass is correctly identified (TP) as many as 326 objects, the seagrass is incorrectly identified (FP) is 6 objects and the seagrass is not identified (FN) is 17 objects. To measure system performance, the calculation process of Precision, Recall, and F1 score is carried out in testing the test dataset. Precision, Recall, and F1 score can be seen in Table 2.
Table 2. Model performance.

| Evaluation Parameter | Value (%) |
|----------------------|-----------|
| Precision            | 98.19     |
| Recall               | 95.04     |
| F1 Score             | 96.58     |

The precision value can show how precise the model determines the positive object that corresponds to the actual [18]. Based on the results of testing on 50 seagrass images, the system's precision in classifying objects accurately reached 98.19%. While recall measures the model's ability to find all positive/daydream objects, based on testing the system can perform a recall of 95.04%. And the F1-Score value is 96.58%. A high percentage indicates that the deep learning model is good [24]. This study only detects one object, so there is no need to look for a true negative (TN) value so that the measurement of the accuracy value cannot be carried out because there is no TN value.

3.4. Percentage of seagrass coverage

The last step in this research is the percentage of seagrass cover area in the test data or before detection and the seagrass coverage area as a result of segmentation. In this process, 50 datasets are used which are also images from underwater camera recordings. The dataset used is different from that used in the training process. The comparison of the percentage results before and after the segmentation process can be seen in Figure 8.

Figure 8. Percentage of seagrass coverage before segmentation process and after Mask R-CNN process.

The graph in Figure 8 is a comparison graph of the percentage of seagrass coverage area obtained before the segmentation process and after the Mask R-CNN process which has segmented the seagrass. As shown in the graph, it is known that the percentage of seagrass coverage obtained has a significant difference. In the calculation of test data or before segmentation, the percentage of seagrass coverage starts from 2.68% to 25.89% while the results after the segmentation process are 2.14% to 14.01%. In Figure 9, it is known that the coefficient of determination for the area of seagrass coverage is 0.6119 with the equation $y = 0.3225x + 2.6257$. The correlation coefficient value of 0.7822 shows a strong relationship with the percentage of seagrass coverage before and after the segmentation results are obtained. The difference in the results obtained can be caused by the results of seagrass segmentation in the validation process, although it has a fairly large detection percentage, the segmented coverage area
is not the whole part of the seagrass. This can be caused by the poor quality of the recorded image so that some parts of the seagrass such as the tip of the seagrass or the small part of the seagrass cannot be segmented perfectly and tend to be wrong.

![Figure 9. Correlation of seagrass coverage in data test image and segmentation image.](image)

4. Conclusion
This research has successfully applied the Mask R-CNN to the Seagrass Enhalus acoroides. The result show a Precision value of 98.19\%, recall of 95.04\%, and F1-Score of 96.58\%. High Precision, Recall and F1-Score indicate that the model is accurate in detecting seagrass correctly, the classifier produced by the model is accurate in detecting required objects and classes can be recognized properly. The correlation value obtained in the segmentation results shows a strong relationship from the data test before and after segmentation. Mask R-CNN is good for generating segmentation on the seagrass object Enhalus acoroides. Based on these results, it can be concluded that the R-CNN Mask can be used to detect seagrass and segment seagrass well.

Acknowledgments
This work is partly supported by IPB Agromaritim Research Grant No. 7824/IT3.L1/PT.01.03/P/B/2021 to Prof. Indra Jaya.

Reference
[1] Rahmawati S, Andri I, Indarto H S and Muhammad H A 2014 Seagrass Monitoring Guide. (Jakarta: CORE-MAP LIPI) p 1
[2] Jesajas D, Raunsay E K, Aiso L E and Dimara L 2016 Analysis of seagrass species in the waters of Yendidori Village Biak Numfor Regency J. Biologi 8 113–30
[3] Hernawan U E, Nurul D M S, Indarto H S, Suyarso, Marindah Y I, Kasih A and Rahmat 2017 Status of Seagrass Indonesia 2017 (Jakarta: Pusat Oseanografi-LIPI) p 12
[4] Hutomo and Nonji 2014 Seagrass Field Monitoring Guide (Jakarta: CRITC COREMAP CTI LIPI) p 2
[5] Bengkal K P, Manembu I S, Sondak C F A, Wagey B T, Schaduw J N W and Lumingas L J L 2019 Identification of seagrass and echinoderm diversity in conservation efforts JITKT 1 29-39
[6] Yusup I M 2020 Automatic recognition of Chaetodontidae family reef fish using deep learning Bachelor Thesis IPB University, Bogor p 2
[7] Cui S, Zhou Y, Wang Y and Zhai L 2020 Fish detection using deep learning Hindawi 2020(3738108) 1-13
[8] LeCun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436-44
[9] Anas D F 2020 Design and implementation of fish freshness detection algorithms using the method deep learning Bachelor Thesis IPB University Bogor p 1
[10] Redmon J and Farhadi A 2017 YOLO9000 better faster stronger In Proceedings of the IEEE conference on computer vision and pattern recognition Oct. 22 2017 to Oct. 29 2017 Venice Italy p 7263-271
[11] Ratnasari Y 2020 Application of deep learning methods to count vannamei shrimp seeds (Litopenaeus vannamei) Bachelor Thesis IPB University Bogor p 2
[12] He K, Gkioxari G, Dollár P and Girshick R 2017 Mask r-cnn Proceedings of the IEEE conference on computer vision and pattern recognition Oct. 22 2017 to Oct. 29 2017 Venice Italy p 2980-88
[13] Hafiz A M and Bhat G M 2020 A survey on instance segmentation: state of the art International journal of multimedia information retrieval 9 171–89
[14] Satria M G A 2019 Design and performance test of Unmanned Surface Vehicle (USV) for mapping water quality and underwater vegetation Master Thesis IPB University Bogor p 27
[15] Johnson J W 2018 Adapting mask-rcnn for automatic nucleus segmentation 1-7
[16] Alfina O 2012 Comparative analysis of backpropagation neural network with simple perceptron in recognizing leaf image Master Thesis Universitas Sumatera Utara North Sumatra p 11
[17] Mahardhika A A, Saptono R and Angraininginsih R 2016 Classification system for customer feedback and recommended solutions for complaints at UPT Puskom UNS with the naive bayes classifier algorithm and cosine similarity Teknol. Inc. ITSmart 4(1) 36-42
[18] Dewi I A, Kristiana L, Darlis A R and Dwiputra R F 2019 Deep learning retinanet based car detection for smart transportation network Elkomika 7 570–84
[19] Saputro I W and Sari B W 2019 Naive bayes algorithm performance test for predicting student study period Citec 6 1-11
[20] Shafira T 2018 Implementation of convolutional neural networks for tomato image classification using hard Bachelor Thesis Universitas Islam Indonesia Yogyakarta p 51
[21] Thohari A and Hertanto G B 2018 Implementation of convolutional neural network for GPU-based motoGP race classification CENTIVE 1(1) 50–55
[22] Wibawa M S 2016 Effect of activation function optimization and number of epochs on artificial neural network performance Sist. and Inform 11 167–74
[23] Santoso S A 2020 Design and implementation of fish identification algorithm with deep learning method Bachelor Thesis IPB University Bogor p 10
[24] Natan O, Gunawan A I and Dewantara B S B 2016 Grid svm aquaculture data processing machine learning applications Elect. Engineering 15 7-17