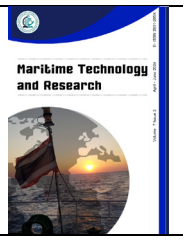
Maritime Technology and Research
https://so04.tci-thaijo.org/index.php/MTR



Research Article

Spectral-Spatial Deep Learning model for seaweed cultivation mapping using PlanetScope imagery in Pangkajene and Islands Regency

Marzuki[1,*], Sanjiwana Arjasakusuma[2], Nurul Khakhim[2], Pramaditya Wicaksono[2], Nur Mohammad Farda[2] and Nur Laila Eka Utami[3]

[1]*Master of Remote Sensing, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Department of Geographic Information Science, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[3]*Bachelor in Cartography and Remote Sensing Study, Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia*

[*]*Corresponding author's e-mail address: marzuki1999@mail.ugm.ac.id*

Article information
Received: July 17, 2024
Revision: November 18, 2024
Accepted: November 22, 2024

Keywords
Spectral-Spatial;
Deep Learning;
Seaweed cultivation;
Remote sensing

Abstract

The efficient mapping of seaweed cultivation over large areas is essential for supporting sustainable management of coastal resources. This study introduces a novel Spectral-Spatial Deep Learning model that integrates spectral and spatial data from high-resolution remote sensing imagery to automate and improve the accuracy of seaweed cultivation mapping. Based on a Convolutional Neural Network architecture, UNet, enhanced with a Spectral-Spatial Attention Module, the model effectively captures the complex relationships between seaweed and its environment. PlanetScope imagery, known for its high spectral and spatial resolution, serves as the primary input data. The model's performance was evaluated using evaluation metrics, achieving an accuracy of 94.71 %, loss of 13.09 %, precision of 80.93 %, recall of 73.63 %, and Intersection over Union (IoU) of 48.51 % on the training data. For the validation data, the model attained an accuracy of 93.64 %, loss of 16.75 %, precision of 84.34 %, recall of 57.57 %, and IoU of 42.98 %. These results demonstrate the model's ability to rapidly and accurately map seaweed cultivation areas, making it a valuable tool for environmental monitoring.

1. Introduction

Seaweed cultivation is a key fishery commodity for Indonesia's coastal communities, playing an important role in supporting the local economy. With vast coastal waters, Indonesia, particularly South Sulawesi Province, holds significant potential for seaweed production, with a cultivation area of up to 250,000 hectares (BPS Provinsi Sulawesi Selatan, 2021). This substantial potential highlights the need for effective monitoring systems to manage seaweed cultivation sustainably.

Mapping large-scale seaweed cultivation areas presents several challenges that require the integration of advanced technologies. Remote sensing, which has been extensively used in environmental and coastal resource monitoring, is particularly well-suited for this task due to its ability to provide efficient, long-term, and cost-effective monitoring across large areas (Jin et al., 2023). In this research, the integration of remote sensing data with artificial intelligence (AI) technology, particularly Deep Learning (DL), is proposed to address the limitations of traditional methods in seaweed mapping.

Conventional methods, such as visual interpretation, work well for small areas, but are time-consuming and require expert knowledge (Langford et al., 2021). Digital classification methods (Andréfouët et al., 2021; Pratama & Albasri, 2021) also face challenges, including spectral confusion between seaweed and other marine features and the need for parameter adjustments with new datasets. Recently, AI technologies like machine learning (ML) have been applied to seaweed mapping (Cheng et al., 2022; Nurdin et al., 2023), but basic ML approaches often rely on manual feature extraction and struggle with large, high-resolution datasets. To overcome these challenges, DL offers a promising solution by automatically extracting complex, non-linear patterns from remote sensing images, which is crucial for mapping heterogeneous objects like seaweed cultivation (LeCun et al., 2015; Ma et al., 2019).

In recent years, DL algorithms have gained popularity in remote sensing image analysis, finding applications in tasks such as image classification, segmentation, and object detection. For seaweed cultivation mapping, DL models can automatically learn complex patterns from image data, eliminating the need for extensive manual feature extraction or domain expertise (Marquez et al., 2022). Among the most widely used DL models for remote sensing data is the Convolutional Neural Network (CNN), which processes data in array form and is particularly suited for high-resolution, multiband remote sensing imagery (LeCun et al., 2015).

One specialized CNN architecture commonly used for image segmentation is U-Net. Originally developed for biomedical imaging (Ronneberger et al., 2015), U-Net has since been adapted for various remote sensing applications. For example, studies have demonstrated its effectiveness in semantic segmentation of high-resolution imagery (Wang et al., 2023), offshore aquaculture extraction (Liu et al., 2023), and aerial building segmentation (Li et al., 2021). The U-Net architecture features a U-shaped design, consisting of a contraction path (encoder) for feature extraction and an expansion path (decoder) for up-sampling and image segmentation (Alam et al., 2021). However, despite its strengths, U-Net has limitations, particularly in accurately identifying object boundaries, as it primarily focuses on spatial features. Remote sensing images also contain valuable spectral features that are essential for tasks like seaweed mapping, which involves distinguishing objects based on spectral variations across different bands (Zhu et al., 2021).

To address these limitations, U-Net architecture was enhanced by incorporating spectral-spatial attention mechanisms. These attention modules allowed the model to focus on both significant spectral features and spatial dependencies, improving its ability to accurately classify objects while reducing redundant information and noise (Mei et al., 2019; Dang et al., 2022). The integration of these mechanisms is crucial for applications such as seaweed cultivation mapping, where high accuracy and precision are required for effective monitoring.

The integration of remote sensing data and DL in mapping seaweed cultivation needs further development to support the provision of spatial data on seaweed farming. One of the remote sensing datasets with significant advantages in terms of resolution is PlanetScope imagery. PlanetScope provides high spatial resolution (Ground Sample Distance of approximately 3 m), enabling detailed analysis of seaweed cultivation areas. Its frequent revisit times, with daily satellite overpasses, allow for continuous monitoring of seaweed cultivation dynamics over time. The PlanetScope spectral bands cover visible and near-infrared regions (**Table 1**), which are crucial for distinguishing seaweed from other marine features based on their spectral characteristics (Planet, 2023).

The complexity of mapping seaweed cultivation, particularly in coastal areas with diverse marine environments, requires advanced techniques capable of capturing intricate spectral and spatial variations. Traditional remote sensing methods for seaweed cultivation mapping often rely on simple enhancement techniques or spectral indices, which may have limitations in accurately distinguishing seaweed from other marine features. These methods can also be sensitive to variations in environmental conditions, such as water clarity. To address these limitations, a Spectral-Spatial Deep Learning model, capable of learning complex and non-linear relationships between spectral and spatial features in remote sensing data, was proposed. PlanetScope imagery was chosen as the primary

data source due to its high spatial resolution, frequent revisit times, and spectral bands well-suited for differentiating seaweed from other marine features. This combination of characteristics provided a valuable dataset for training and evaluating the model that was developed. Therefore, this approach was expected to enhance the accuracy and robustness of seaweed detection, even in challenging environments.

2. Research method

2.1 Research location

This research was conducted in Pangkajene and Islands Regency (Pangkep), which focuses on the Marang District area. Geographically, Marang District is located between 119º29'24'' - 119º34'00'' East (E) and between 4º37'12''- 4º2'33'' South (S). The map of this research area can be seen in **Figure 1**. This area is one of the areas in Pangkep Regency where people actively produce seaweed.

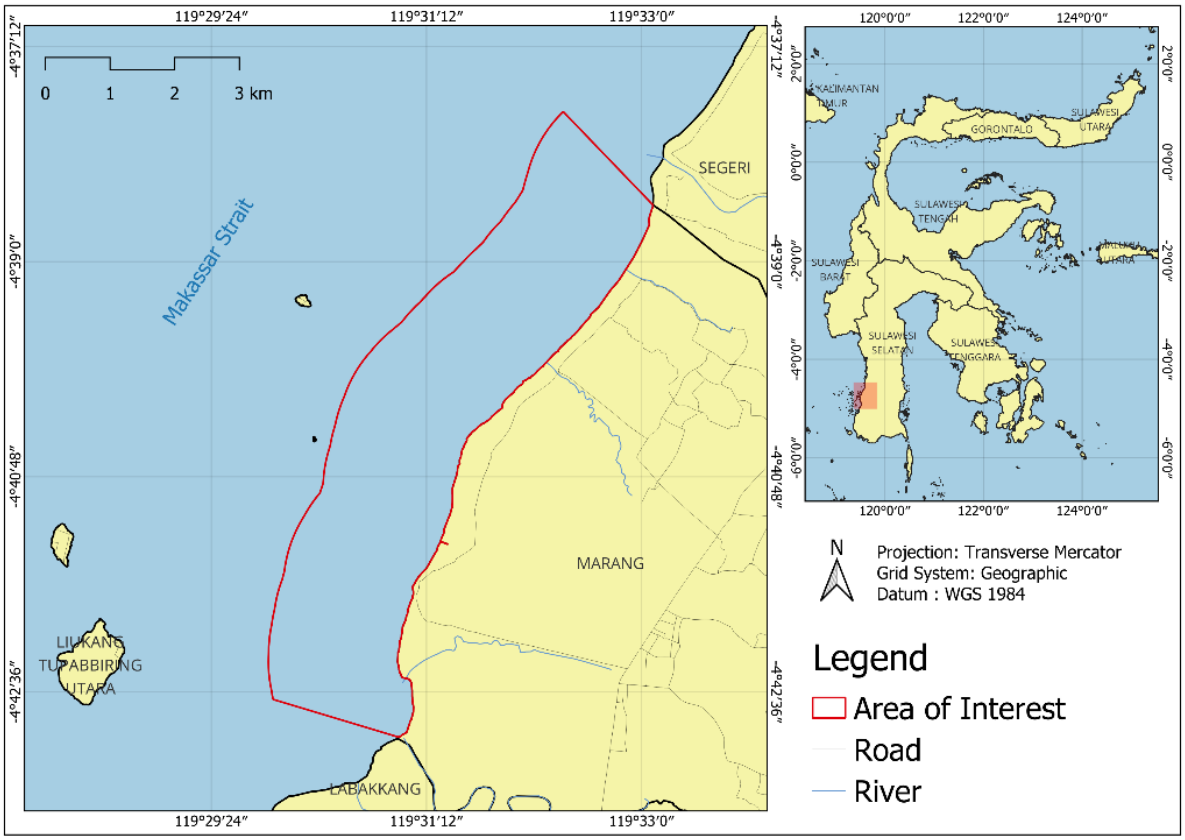


Figure 1 Map of the research area.

2.2 Input data description

The data source used in this study was remote sensing imagery from PlanetScope. PlanetScope is a satellite constellation owned by Planet, consisting of multiple individual satellite groups. The latest product available from the PlanetScope satellites is the third generation, known as SuperDove (PSB.SD). This data can be accessed at https://www.planet.com. The SuperDove sensor is equipped with eight spectral bands, including coastal blue, blue, green I, green II, yellow, red, red-edge, and near-infrared (NIR). The specific wavelength range for each band can be found in **Table 1**. PlanetScope imagery comes in several levels, each representing different stages of correction. PlanetScope Level 3B is a PlanetScope Ortho Scene product that has undergone orthorectification and radiometric correction, either to Top of Atmospheric Radiance (at sensor) or Surface Reflectance (SR). The Ground Sample Distance (GSD) of PlanetScope imagery is 3.7 m, but orthorectified

imagery offers a spatial resolution of 3 m. Additionally, PlanetScope provides daily temporal resolution, making it highly suitable for continuous monitoring purposes (Planet, 2023).

The PlanetScope imagery used as input data to build the model was acquired on March 22, 2021, May 1, 2021, and May 12, 2021. The imagery was selected from the months of March to May because, visually, these months exhibit significant seaweed cultivation activity. Additionally, the chosen scenes have high image quality, with cloud cover below 20 % and full coverage of the entire study area. The imagery from April 19, 2023 was used for prediction purposes. All these images had the same correction level in order to maintain consistency across the data. In this process, we acquired selected PlanetScope imagery with the level of correction in the SR level 8-band analytic images and harmonized images. However, the NIR channel was not included in the entire process of the model, because it dis not provide significant information on seaweed, resulting in the input only having 7 bands. All processes in this study were carried out on a computer with the following specifications: 1) Intel i9-14900F-20 processor, and 2) MSI RTX 4080 16GB Ventus 3X OC GPU. The Deep Learning model was built using the TensorFlow library.

Table 1 Spectral bands of the SuperDove (PSB.SD) sensor of the PlanetScope imagery.

Bands	Wavelengths (nm)
Coastal Blue	431 - 452
Blue	465 - 515
Green I	513 - 549
Green II	547 - 583
Yellow	600 - 620
Red	650 - 680
Red-Edge	697 - 713
NIR	845 - 885

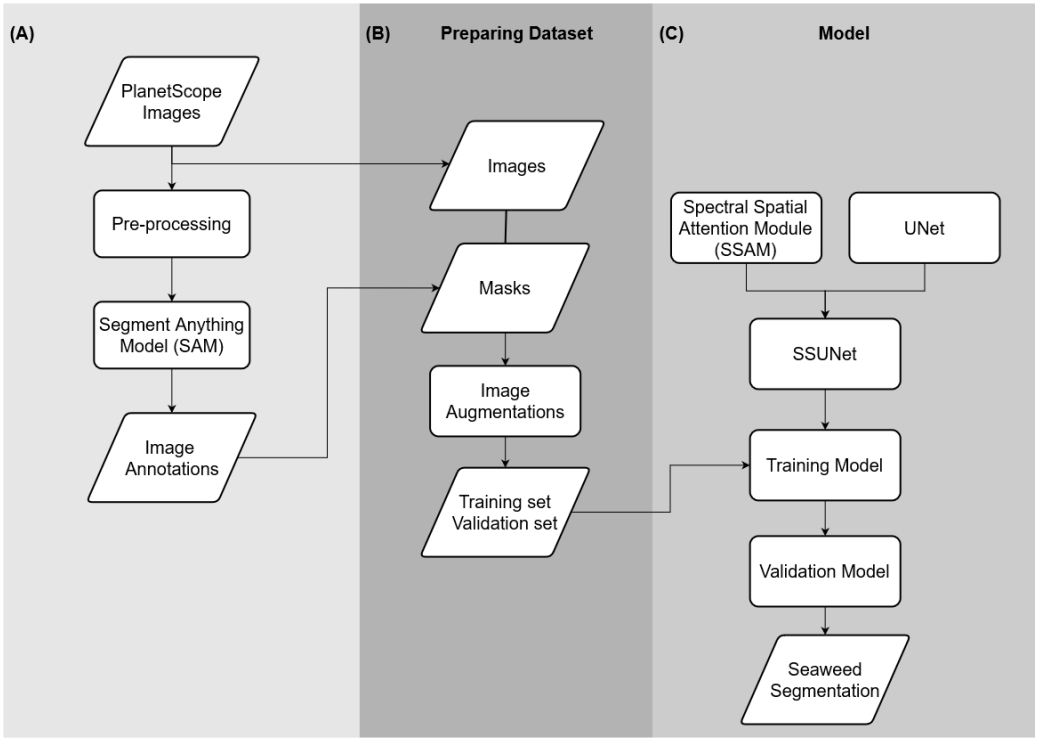


Figure 2 Research flowchart of mapping seaweed using DL.

2.3 Method

The method used in this study can be split into three sections (**Figure 2**): (A) Building training data from preprocessing of PlanetScope imagery and generation of image annotations, (B) generation of datasets, and (C) seaweed mapping and accuracy assessment.

2.3.1 Building training data

2.3.1.1 Data pre-processing

The purpose of pre-processing data is to visually enhance the image quality before further processing to produce image annotations. The processes involved: 1) image sharpening, using contrast stretching algorithms, and 2) applying spatial or focal filtering with a moving window, to obtain minimum value of the filtered pixel values (Bajpai et al., 2017).

In detail, the first process involves sharpening the image by enhancing the contrast linearly. The minimum and maximum values of the image are set to a new specified range, utilizing the full range of available brightness values. This process stretches the pixel values between two specific quantiles, effectively improving the visual contrast. This can be referred to as the min-max contrast linear stretch, as described in Eq. (1). The stretched pixel value (R'_{ij}) at position (i,j) is computed using:

$$R'_{ij} = \left(\frac{R_{ij} - R_{min}}{R_{max} - R_{min}}\right) * \ Number\ of\ Intensity\ Levels \quad (1)$$

In this equation, R_{ij} is the original pixel value at position (i,j), R_{min} and R_{max} represent the minimum and maximum values in the original image, respectively, and the "Number of Intensity Levels" defines the total number of possible intensity levels assignable to a pixel. For instance, in a grayscale image with a maximum intensity of 255, this makes 255 the highest intensity value used.

The second process involves focal filtering using a 3×3 kernel applied to each pixel. This kernel defines the neighborhood of each pixel considered in the focal filtering operation. The minimum function is applied to extract the minimum value from all the pixels within the 3×3 window surrounding each central pixel. This results in new data, where each pixel contains the minimum value from the surrounding 3×3 window. The object boundary become smoother, so that the object's appearance becomes more real. Mathematically, the focal filtering process can be calculated as shown in Eq. (2).

$$F'_{ij} = min\{F_{i+k,j+l} | -1 \leq k \leq 1, -1 \leq l \leq 1\} \quad (2)$$

In this equation, F'_{ij} represents the pixel value in the filtered image at position (i,j) after applying the filter. $F_{i+k,j+l}$ corresponds to the pixel values from the original image that fall within a 3×3 window centered at pixel (i,j). The offsets k and l range over -1, 0, 1, thereby including all neighboring pixels within the window.

The outcome of these processes was expected to yield a more accurate seaweed masking for the training data. Subsequently, the output imagery was adjusted to match the specification configurations of the next process, where the images were scaled to 8-bit (0 - 255) and reduced to only 3 bands (Red, Green I, and Blue bands) which were then composited into RGB composite or a true color image.

2.3.1.2 Segment anything model

Since the collection of label data through the image annotation process needed to be quick and accurate, the Segment Anything Model (SAM), developed by Meta AI (Kirillov et al., 2023), was used; a state-of-the-art zero-shot segmentation process to segment objects in images. SAM has the capability to generalize various datasets for image segmentation without requiring additional training

data (Osco et al., 2023). Therefore, this method was used to automatically obtain labels for seaweed and non-seaweed. The prepared and enhanced PlanetScope imagery data were used as input data, and SAM was applied to this data using a transfer learning system. Principally, SAM employs a zero-shot learning concept by utilizing prompting. However, this research did not use specific prompting, instead utilizing the SAM automatic mask generator to produce masks (Wu & Osco, 2023). The use of the SAM's automatic mask generator can directly produce labels; however, to ensure these labels were accurate, it was necessary to configure its advanced parameters such as points per side, Intersection over Union (IoU) predict threshold, stability score threshold, and others. Furthermore, the results of SAM were used as label data for the next process.

2.3.2 Preparing datasets

2.3.2.1 Image augmentations

Before the image augmentation process, the images and their annotations were first cropped into several patches according to the input dimension size to be used. Then, image augmentation is performed on these patches. The purpose of image augmentation is to increase the number of new datasets and enhance image variability to prevent overfitting. The augmentation methods used to enhance image variability include center crop, random rotate, random sized crop, horizontal and vertical flip, and sharpen (Takahashi et al., 2015). Image augmentations were executed using the Albumentation library.

2.3.2.2 Split data

The models were built from the collected dataset, which comprised 2,835 pairs of images and masks. This dataset was derived from various image scenes representing seaweed cultivation water conditions and had undergone image augmentation. The data was then divided into two groups: training data and validation data, with proportions of 80 and 20 %, respectively.

2.3.3 Training the Deep Learning model

The model training process involved constructing a U-Net architecture, which integrated a combination of a Spectral Attention Module (SpeAM) and a Spatial Attention Module (SpaAM). The combination of these two modules led to the proposed Spectral-Spatial Attention Module (SSAM). Originally, SSAM was derived from the Convolutional Block Attention Module (CBAM), which was proposed by Woo et al. (2018). While CBAM is predominantly used in non-remote sensing data, several studies have adapted it for remote sensing data. SSAM can be integrated into various CNN architectures, including UNet. Studies that have incorporated SSAM into UNet have demonstrated strong model performance (Dang et al., 2022; Wang et al., 2022; Zhu et al., 2021). In this study, the SSAM configuration from the work of Yan et al. (2021) was adopted, whereby SSAM was placed in each layer of the UNet to preserve spectral and spatial features across all layers. However, several modifications were made, including adjusting the input image size, fine-tuning hyperparameters, and updating the evaluation metrics used. These adjustments were made to achieve an optimal configuration and enhance the model's performance in mapping seaweed cultivation. A more detailed explanation of the modules used and the overall architecture are outlined below:

2.3.3.1 Spectral Attention Module (SpeAM)

The SpeAM was used to extract spectral features present in each channel of the PlanetScope imagery. As illustrated in **Figure 3**, the input feature of SpeAM is represented by $F \in \mathbb{R}^{H \times W \times C}$, where H is the height, W is the width, and C is the number of channels. The input data underwent global average pooling (AvgPool) and global max pooling (MaxPool) operations. These operations each produced a 1*1*C vector. AvgPool can be calculated based on Eq. (3) and MaxPool on Eq. (4). The average value of each channel spectral denotes $F_{avg} \in \mathbb{R}^{1 \times 1 \times C}$ and the maximum value of each channel spectral denotes $F_{max} \in \mathbb{R}^{1 \times 1 \times C}$.

$$F_{avg}(c) = \frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=i}^{W} F\,(i,j,c) \quad \forall c \in \{1,2,\dots,C\} \tag{3}$$

$$F_{max}(c) = \max_{i,j} F\,(i,j,c) \quad \forall c \in \{1,2,\dots,C\} \tag{4}$$

The resulting F_{avg} and F_{max} vectors were then combined and processed through a network consisting of a multilayer perceptron (MLP) with a single hidden layer. Through, these two vectors passed through two Fully Connected Layers (FCL), followed by the Rectified Linear Unit (ReLU) activation function (σ). After each spectral vector passed through the shared network or MLP, two output vectors were produced, namely the MLP output of F_{avg} (M_{avg}) and the MLP output of F_{max} (M_{max}). Both can be calculated based on Eqs. (5) and (6), respectively. In the calculation, there is a weight ($W \in \mathbb{R}^{C\times\frac{C}{r}}$) of each FCL which has a reduction ratio (r). Each weight represents W_1 for the first FCL and W_2 for the second FCL.

$$M_{avg} = W_2\sigma(W_1F_{avg}) \tag{5}$$

$$M_{max} = W_2\sigma(W_1F_{max}) \tag{6}$$

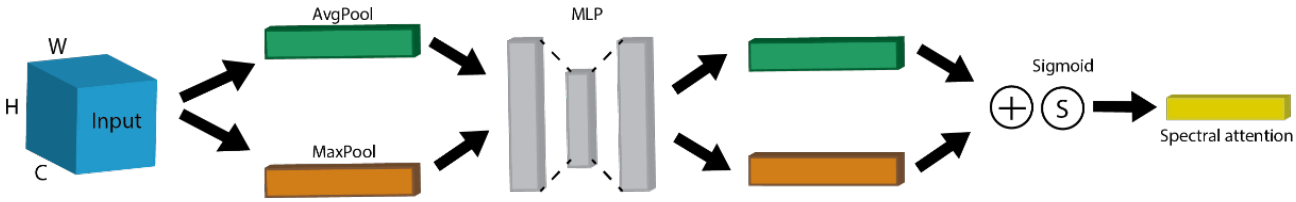
These output vectors were then combined using element-wise addition (+). Finally, the spectral attention weights were obtained using the sigmoid activation function (δ). This can be calculated based on Eq. (7):

$$M_c = \delta(M_{avg} + M_{max}) \tag{7}$$

The final result of the spectral attention map calculation ($M_c \in \mathbb{R}^{1\times1\times C}$) was then multiplied by the input data to produce spectral attention (SpeAM) output (Dang et al., 2022; Wang et al., 2022; Yan et al., 2021; Zhao et al., 2024; Zhu et al., 2021). SpeAM can be calculated based on Eq. (8):

$$F' = F * M_c \tag{8}$$

In this equation, * represents the element-wise multiplication, and $F' \in \mathbb{R}^{H\times W\times C}$ represents the output of the SpeAM.



2.3.3.2 Spatial Attention Module (SpaAM)

The Spatial Attention Module (SpaAM) was used to effectively learn the spatial features of seaweed in PlanetScope imagery. As illustrated in **Figure 4**, the input feature of SpaAM is represented by the result of SpeAM (F'). The process began by applying AvgPool (F'_{avg}) and MaxPool (F'_{max}) operations to the input data along the channel axis of the PlanetScope imagery. This resulted in an output of Height (H) * Width (W) * 1 or $F'_{avg/max} \in \mathbb{R}^{H\times W\times1}$. Mathematically, AvgPool can be calculated based on Eq. (9) and MaxPool based on Eq. (10):

$$F'_{avg}(i,j) = \frac{1}{C}\sum_{k=1}^{C} F'(i,j,k) \quad \forall i \in \{1,\dots,H\}, j \in \{1,\dots,W\} \tag{9}$$

$$F'_{max}(i,j) = \max_{k} F'(i,j,k) \quad \forall i \in \{1,\dots,H\}, j \in \{1,\dots,W\} \tag{10}$$

Next, the outputs from these two operations (F'_{avg} and F'_{max}) were concatenated ($concat$), followed by a convolution operation ($Conv2D$) and a sigmoid activation function (δ) to obtain the spatial attention output, which represents the contribution of each pixel (Dang et al., 2022; Wang et al., 2022; Yan et al., 2021; Zhao et al., 2024; Zhu et al., 2021). Mathematically, SpaAM can be calculated based on Eqs. (11), (12), (13), sequentially:

$$F" = concat(F'_{avg,}F'_{max,}axis = -1) \tag{11}$$

$$M_s = \delta(Conv2D(F", filters = 1, kernel\_size = 7)) \tag{12}$$

$$F''' = F' * M_s \tag{13}$$

In these equations, $F" \in \mathbb{R}^{H\times W\times 2}$ denotes the result of concatenated the average and max pooled features along the channel spectral axis, $M_s \in \mathbb{R}^{H\times W\times 1}$ denotes the spatial attention map, $Conv2D$ denotes the convolution operation with 7*7 kernel, and $F''' \in \mathbb{R}^{H\times W\times C}$ denotes the output of the SpaAM.

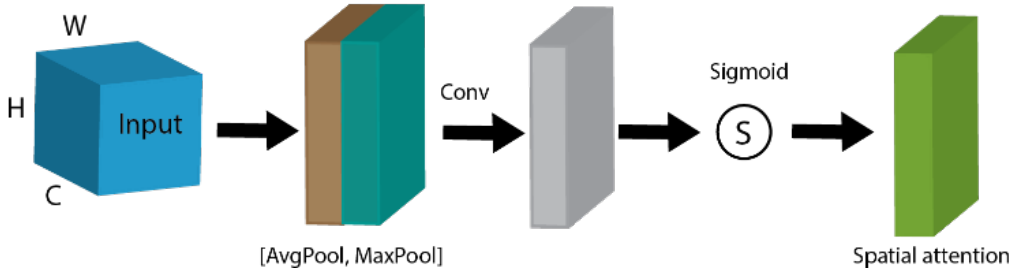


Figure 4 Spatial Attention Module (SpaAM).

2.3.3.3 Spectral-Spatial attention module (SSAM)

The Spectral-Spatial Attention Module (SSAM) was arranged sequentially to leverage the potential of both spectral and spatial attention simultaneously. SSAM could be integrated into the training process. As illustrated in **Figure 5**, the process began from the image being multiplied by the SpeAM calculation. The result of the SpeAM calculation served as the input for the SpaAM calculation. Then, the input was multiplied by SpaAM. Thus, this sequential operation resulted in SSAM output. Mathematically, SSAM can be calculated based on Eq. (14):

$$F_{SSAM} = F''' = (F * M_c) * M_s \tag{14}$$

In this equation, $F_{CBAM} \in \mathbb{R}^{H\times W\times C}$ was the final output after applying both SpeAM and SpaAM to the input feature map (F).

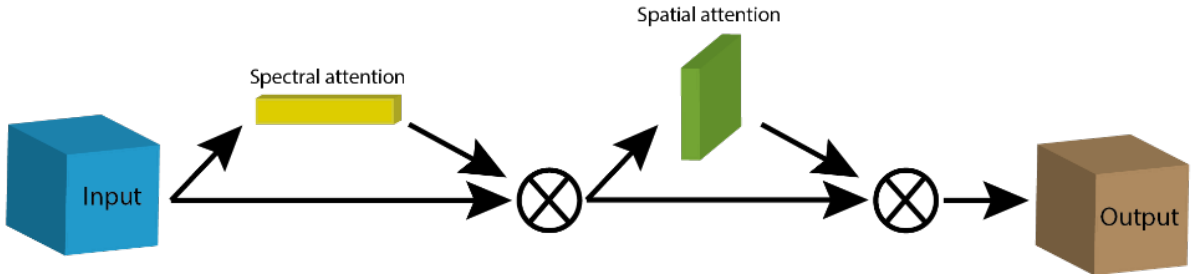


Figure 5 Spectral-Spatial Attention Module (SSAM).

Next, the SSAM was integrated into the U-Net architecture, proposing the Spectral-Spatial attention U-Net (SSUNet) model. This integration was proposed because convolutional layers typically ignore spatial and spectral channel correlations, instead focusing on processing spatial information around the central pixel of the sample.

2.3.3.4 Spectral-Spatial attention UNet (SSUNet) for seaweed mapping

As illustrated in **Figure 6**, the SSUNet architecture broadly consisted of two layers: the encoder and the decoder. In the encoder layer, the process included convolutional layers, pooling, and SSAM. The convolutional layers used 3*3 kernel combined with the ReLU activation function. The convolution process was performed twice, and the results were then applied to the SSAM. Following this, Max pooling with a 2*2 kernel was performed to reduce the dimensions of the feature map.

The decoder layer included transposed convolution (up-convolution), regular convolution, concatenation, and SSAM. The transposed convolution used a 2*2 filter. The output was then concatenated with features from the encoder at the same level in a process called a skip connection. This was followed by two 3*3 regular convolutions, each accompanied by the ReLU activation function. The results of these convolutions were then processed by the SSAM before undergoing up-convolution to the next layer. Finally, the output from the decoder layer passed through a 1*1 convolution with sigmoid activation. The sigmoid activation function provided a segmentation mask that represented pixel-by-pixel classification of the image. SSAM was placed together in each layer of the U-Net (Yan et al., 2021). Hence, the SSUNet was proposed in this study to be used for mapping seaweed cultivation areas.

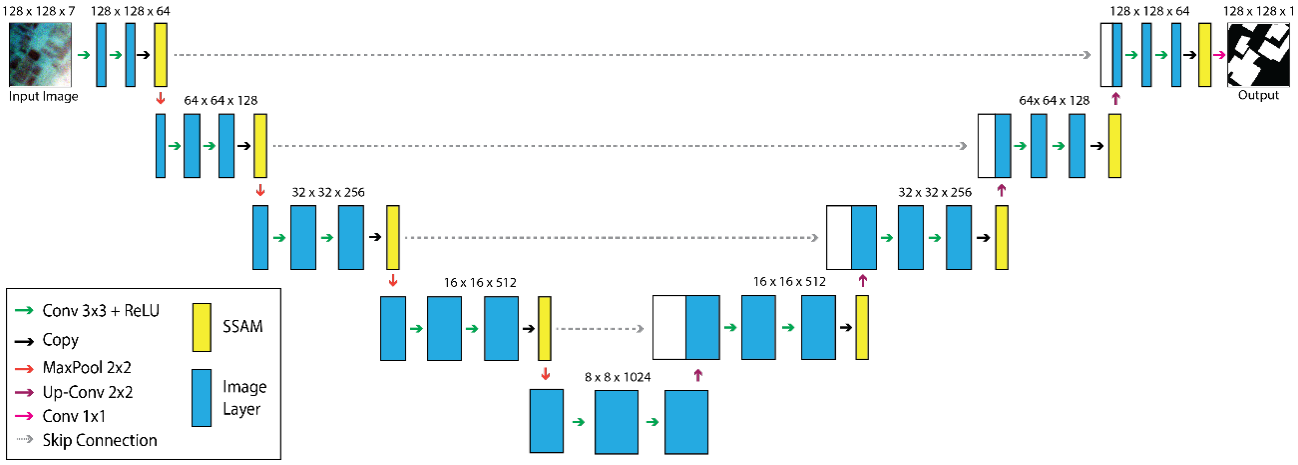


Figure 6 Spectral-spatial attention UNet (SSUNet).

Before training the model, the hyperparameters were carefully configured to produce the best model. The hyperparameters used in the training process included the learning rate, batch size, and the number of epochs. The model was trained using the Adam optimizer and incorporated an early stopping mechanism to prevent overfitting.

2.3.4 Evaluation metrics

During the training process, the model required performance evaluation or accuracy testing to achieve an optimal model and avoid overfitting. The data used in the training process included training data and validation data, both of which were run sequentially in each iteration (epoch). The model's performance was evaluated using various metrics, providing accuracy values for each dataset used. The evaluation metrics in this study included accuracy, loss function, precision, recall, and Intersection over Union. Since the segmentation results produced two classes, the loss function used was binary cross-entropy (BCE). In the process of calculating the evaluation metric, the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values were determined. Mathematically, evaluation metrics such as accuracy, precision, recall, and IoU can be calculated using Eqs. (15), (16), (17), and (18), sequentially. The detailed formulations are explained as follows:

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{15}$$

$$precision = \frac{TP}{TP+FP} \tag{16}$$

$$recall = \frac{TP}{TP+FN} \tag{17}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{18}$$

3. Results and discussion

3.1 Building training data

The training data was sourced from PlanetScope imagery, which served as the image input and reference for generating labels. To produce the labels, several steps were applied, including data pre-processing and the Segment Anything Model (SAM). These steps, in more detail, yielded the following outputs:

3.1.1 Data pre-processing

Before pre-processing, the PlanetScope imagery displayed seaweed cultivation areas with low color intensity (**Figure 7**). The seaweed plots appeared dark but blended with the surrounding water. In various experiments conducted, the Segment Anything Model (SAM) struggled to differentiate seaweed from seawater effectively. The pre-processing stage helped resolve this issue by enhancing the contrast of the seaweed cultivation plots, making the boundaries between objects clearer. As shown in **Figure 7**, the pre-processed data reveals that the seaweed plots can now be visually separated. Therefore, this step was crucial, as it significantly improved the quality of SAM segmentation results for seaweed.

3.1.2 Segment Anything Model (SAM)

The application of SAM was expected to produce label data quickly and accurately. Accurate labels were crucial for the model to learn the structure of objects in the imagery and to provide an objective evaluation of the model's performance. Without accurate labels, the model would struggle to identify the location and boundaries of seaweed objects in the images. In a study

by Karimi et al. (2020), it was demonstrated that the presence of noisy labels can severely impact the performance of DL models. Models trained on datasets with a high degree of label noise tend to memorize the training data, rather than learning patterns that can be generalized. As a result, the models perform poorly on new, unseen data because they struggle to effectively distinguish between correct and incorrect labels. Human annotators often introduce bias or errors during the data labeling process, leading to misleading training signals for Deep Learning models. Therefore, in this study, SAM was applied to automatically generate labels for seaweed and non-seaweed areas, aiming to produce labels with high accuracy and reduce the impact of human-induced label noise.

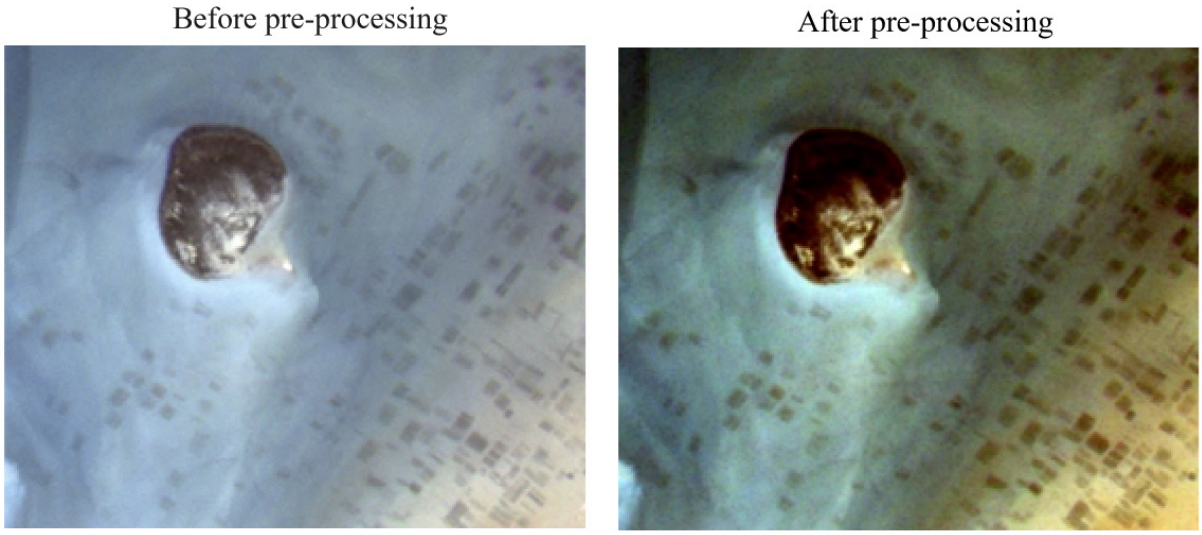


Figure 7 PlanetScope imagery in May 1, 2021, before pre-processing and after pre-processing.

The input image for SAM was a preprocessed image. It is important to note that SAM outputs are instance segmentation, where each object is distinguished even if it belongs to the same class. Therefore, it was necessary to merge objects of the same class and separate those that were not needed. This process resulted in two binary classes: seaweed and non-seaweed. These annotated results were used as labels for the model's training and testing process. Visually, the comparison between the input image, the SAM results, and the labels or mask to be used can be seen in **Figure 8**.

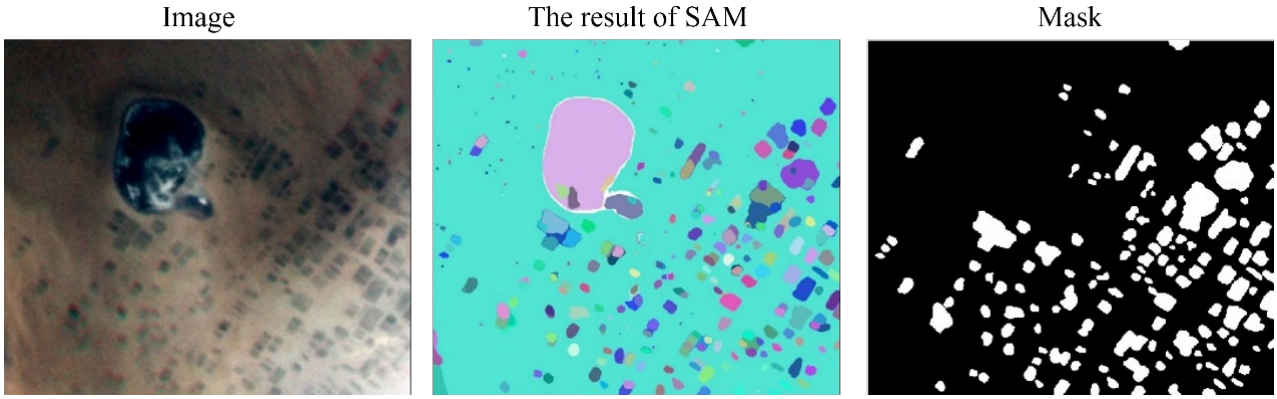


Figure 8 Comparison between the preprocessed image, the SAM results, and the labels or mask on the PlanetScope image dated March 22, 2021.

3.2 Training model

The model training process was based on the designed SSUNet architecture. The input data required for training had to have a uniform and consistent size. In this study, the image size used was 128×128 pixels. A total of 405 pairs of images and labels were collected. However, a small amount of data for building a DL model could have led to the risk of overfitting. Data augmentation is a solution to increase the amount of data by applying transformations to the input data that preserve the shape of the objects. Various data augmentation strategies enhance the variability of the data, enabling the model to better understand the different variations present in the data. As a result of augmentation, the number of data increased to 2,835 pairs of images and labels.

The model training process required setting hyperparameters in a way that yielded the best model. Some of the hyperparameters used in the training process included learning rate, batch size, and the number of epochs. These hyperparameters were uniformly set, with the learning rate set to 0.0001 and adaptively updated during the training process, the batch size set to 32, and the number of epochs set to 100. This configuration was the most optimal combination obtained during the model training trials.

3.3 Accuracy assessments

The best-performing model was obtained after several training trials. The model's performance evaluation could be assessed from the values of the evaluation metrics. The application of the SSUNet architecture in the model training process resulted in accuracy of 94.71 %, loss of 13.09 %, precision of 80.93 %, recall of 73.63 %, and IoU of 48.51 % on the training data. For the validation data, the results showed accuracy of 93.64 %, loss of 16.75 %, precision of 84.34 %, recall of 57.57 %, and IoU of 42.98 %. The comparison of evaluation metric values between the training and validation data during the training process can be seen in **Figure 9**. It was necessary to compare the evaluation metric values obtained during the model training process with those from the validation process. This comparison was crucial for detecting overfitting or underfitting in the model. The performance of evaluation metric values for both training and validation data should have been consistent, with no significant disparity between them. If the values were consistent, it indicated that the model had learned well and was not significantly overfitting or underfitting. The error for the model on the training data reduced, as well as the error on the validation dataset. When a model's performance metrics on both training and validation datasets are similar, it suggests that the model generalizes well to new, unseen data (Colliot, 2023).

Based on the evaluation metrics, the high accuracy on both training and validation data indicated that the model accurately classified the majority of seaweed pixels. The lower loss on training data compared to validation data suggested effective learning of training patterns. The higher precision on validation data implied that the model's positive pixel predictions were highly accurate, with few false positives. Conversely, the lower recall on validation data indicated the model's reduced effectiveness in identifying all positive pixels, leading to many false negatives. The lower IoU on validation data highlighted the model's difficulty in generating seaweed segmentation predictions that precisely matched the actual areas, revealing imperfections in the overlap between predictions and ground truth, especially on previously unseen data.

3.4 Effectiveness of SSUNet for seaweed mapping

To further investigate the effectiveness of the SSUNet model for mapping seaweed cultivation, the model was applied to previously unseen data. The model was tested on several validation datasets to visually assess its ability to accurately predict seaweed cultivation. The spectral and spatial modules were designed to guide the model's focus on the spectral and spatial features present in PlanetScope imagery, which is rich in both spectral and spatial information.

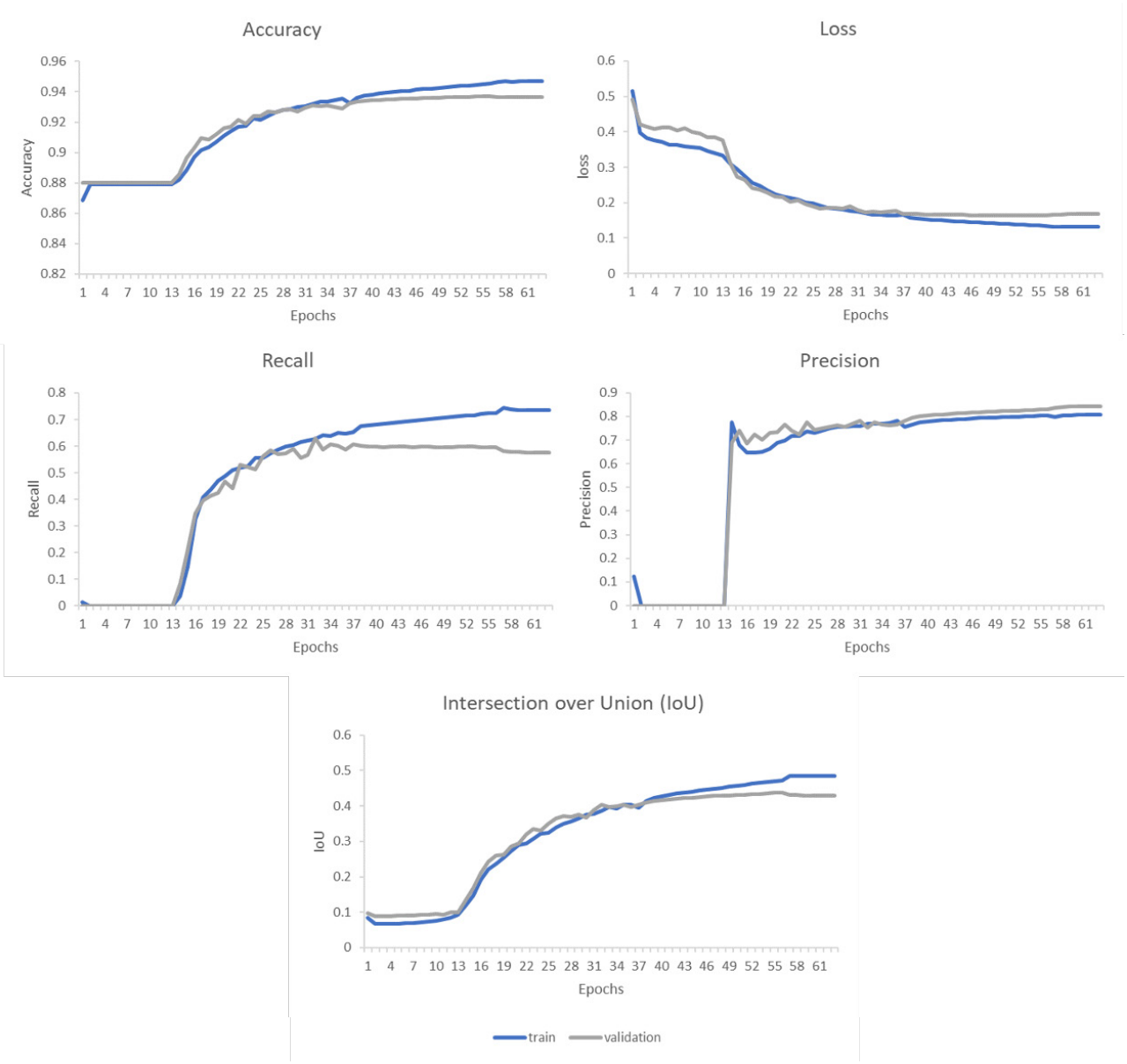


Figure 9 Comparison graph of evaluation metrics: accuracy, loss, precision, recall, and IoU on training and validation data.

The prediction results of the model can be seen in **Figure 10**. Visually, when comparing the image, mask, and prediction, it was evident that the SSUNet model effectively predicted seaweed cultivation areas. The segmentation results clearly represented the patterns of the seaweed cultivation areas, although some boundaries were still imprecise. This imprecision was due to the model's performance imperfections, resulting in some inaccurate predictions. The red boxes in **Figure 10** highlight the discrepancies between the labels and predictions. The predicted object boundaries still exhibited some misclassifications. However, overall, the model accurately identified the features and patterns of seaweed.

Upon closer examination, it is apparent that the dark-toned seaweed cultivation areas in the PlanetScope imagery were predicted accurately. This indicates that the seaweed in these areas has been densely planted and has grown robustly. However, seaweed cultivation areas with lighter tones were not predicted as accurately. This suggests that these areas have not been fully planted, or potentially have not been planted at all, leaving only the stakes of the seaweed cultivation framework visible.

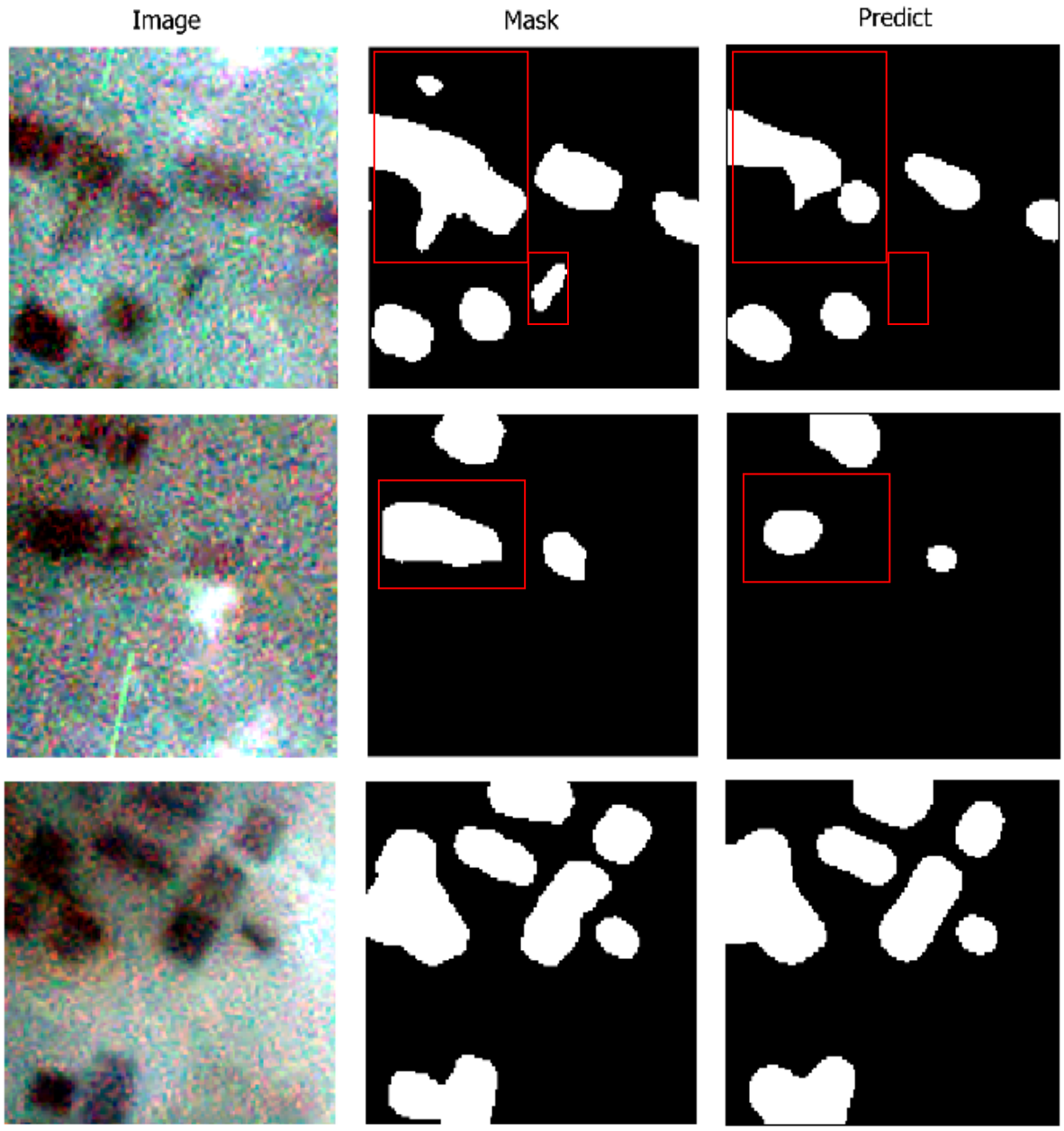


Figure 10 Comparison between image, mask, and prediction.

The SSUNet model can be implemented for multitemporal mapping of seaweed farming areas. Given that PlanetScope imagery has a daily temporal resolution, it holds significant potential for sustainable monitoring of seaweed farming. However, at this stage, the model has only been applied to a single scene from PlanetScope imagery, recorded on April 19, 2023. The model application results are shown in **Figure 11**. The model's prediction results for the image scene were used to extract the seaweed farming area. On that particular date, the estimated seaweed farming area was 383.69 hectares. Meanwhile, the seaweed farming area derived from visual interpretation (ground truth) was determined to cover 433 hectares. Comparing the predicted area to the ground truth reveals a discrepancy of 49.31 hectares. This indicates that the model's predictions still experience some misclassifications, contributing to the observed difference in area. The model's performance, which has not yet been fully optimized, also plays a role in this outcome.

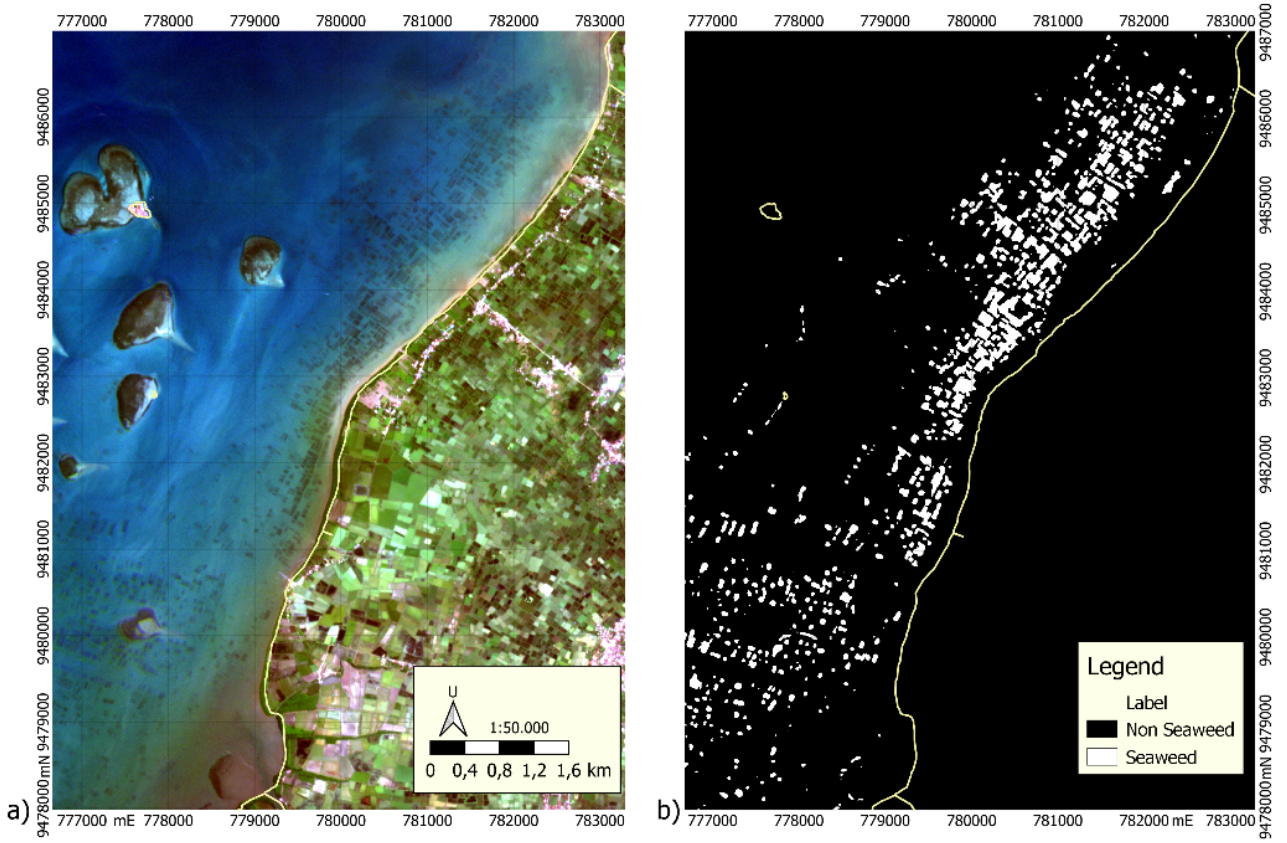


Figure 11 (a) Full scene true color composite PlanetScope image on April 19, 2023, and (b) predicted seaweed cultivations.

A comparison between the true color composite of PlanetScope imagery (**Figure 11a**) and the predicted seaweed farming areas (**Figure 11b**) demonstrates that the model effectively predicted the abundant seaweed farming plots in the water. However, some areas experienced misclassification, such as in the sandbar. Some non-seaweed objects, such as seagrass or coral, were still classified as seaweed. This is because seaweed is cultivated on the water's surface, and in several seaweed farming areas, the substrate consists of seagrass or coral. These underlying objects often blend together in the imagery, making it difficult to separate them both spectrally and spatially. Consequently, distinguishing between seaweed and these other marine features posed a significant challenge for accurate classification.

The complexity of seaweed as an object also presented significant challenges. When visually examined in the imagery, seaweed farming plots can be spatially distinguished based on their planting patterns (Langford et al., 2021). However, spectrally, seaweed is difficult to differentiate from other marine vegetation, such as seagrass and other types of macroalgae. These plants share similar spectral signatures due to the presence of chlorophyll and other pigments, with the primary distinction lying in the intensity of their reflectance values (Dora et al., 2024). In addition to the complexity of seaweed itself, the inherent complexity of PlanetScope imagery as the input data, which exhibits high spectral variability, also contributed to the challenges. This spectral diversity can potentially introduce a significant amount of noise (Kington & Collison, 2022), which made it difficult for the model to accurately detect object boundaries. These complexities in both the object and data significantly impacted the model's ability to perform precise segmentation. Therefore, the model's performance limitations were partly influenced by these factors.

The satisfactory performance of the seaweed cultivation prediction model was largely attributed to the architecture of SSUNet. The integration of the Spectral-Spatial Attention Module (SSAM) in each UNet layer effectively enhanced the representation of spectral information and measured the spatial significance of seaweed features at each model layer. Specifically, the Spectral Attention Module focused on correlations between adjacent spectral dimensions, leading to improved classification outcomes. Meanwhile, the Spatial Attention Module emphasized spatial dependencies by assigning attention weights to each pixel, aiding in capturing local spatial patterns and enhancing overall classification performance (Wang et al., 2022). Consequently, the sequential mechanism of SSAM enabled each layer to adaptively adjust the feature map weights and extract more informative features.

3.5 Future research

The SSUNet model developed for mapping seaweed cultivation still had several limitations, including some inaccuracies in segmentation that could be improved. This study recommends enhancing the quality of the DL-based seaweed cultivation mapping model through refinement techniques, such as post-processing, or training enhancements, like fine-tuning hyperparameters. These steps aim to improve the delineation of object boundaries. Additionally, further exploration of different types or combinations of attention mechanisms is necessary to produce a more optimal and more accurate seaweed cultivation mapping model.

This research aims to serve as a reference for the future development of DL-based seaweed cultivation mapping methods. Various aspects such as environmental parameters, water quality, and oceanographic conditions can be integrated if seaweed cultivation mapping is quickly and easily accessible. This integration will enable more complex analyses for monitoring seaweed cultivation. In addition, the developed model can be applied to various applications for the advancement of seaweed cultivation, such as monitoring seaweed cultivation patterns, extracting the cultivated area that can be validated using field data, and estimating seaweed production. Therefore, the models produced could be used by various stakeholders, including the government and private sectors. This will support the acceleration of seaweed cultivation data production, which will positively impact resource sustainability and strengthen food security, particularly in the field of fisheries resources.

4. Conclusions

In remote sensing imagery, spectral and spatial features can be extracted. PlanetScope imagery, with its high spectral and spatial resolution, holds significant potential for use as input data. The Spectral-Spatial Attention UNet (SSUNet) architecture was implemented for seaweed cultivation segmentation. The addition of the spectral-spatial attention module (SSAM) to the UNet enhanced the extraction of seaweed's spectral and spatial features. The model's training process was evaluated using metrics such as accuracy, loss, precision, recall, and Intersection over Union (IoU). The results indicated that the model achieved high accuracy values (94.71 and 93.64 %) and relatively low loss values (13.09 and 16.75 %) for the training and validation datasets. These outcomes suggest that the model performs consistently, though there is a slight overfitting tendency. The high precision values for both datasets (80.93 and 84.34 %) demonstrate that the model can accurately identify positive pixels, but the lower recall values (73.63 and 57.57 %) indicate challenges in detecting all positive pixels. As a result, the IoU values (48.51 and 42.98 %) are moderate, suggesting that, although the model produces correct predictions, there remains room for improvement in the overlap between predictions and ground truth. Overall, the evaluation metrics demonstrate that, while the model detects most objects, inaccuracies in boundary prediction result in suboptimal segmentation. Nevertheless, the model performs well in rapidly and accurately extracting seaweed cultivation areas. The developed seaweed cultivation mapping model was applied to one of the image scenes in the study area. The prediction results revealed a clear distribution of seaweed cultivation areas, leading

to an estimated area of 383.96 hectares. This highlights the significant potential for further seaweed farming development.

Acknowledgments

The authors would like to thank the support of the domestic cooperation unit and the Remote Sensing Laboratory, Faculty of Geography, Gadjah Mada University, and the anonymous reviewers for helping to improve this manuscript. We would like to thank Planet Team for the opportunity to use and assess the performance of PlanetScope images. This publication has been funded by the Department of Foreign Affairs and Trade Australia through KONEKSI. The views expressed in this publication are the authors' alone and are not necessarily the views of the Australian Government.

References

Alam, M., Wang, J. F., Guangpei, C., Yunrong, L., & Chen, Y. (2021). Convolutional neural network for the semantic segmentation of remote sensing images. *Mobile Networks and Applications*, *26*(1), 200-215. https://doi.org/10.1007/s11036-020-01703-3

Andréfouët, S., Dewantama, I. M. I., & Ampou, E. E. (2021). Seaweed farming collapse and fast changing socio-ecosystems exacerbated by tourism and natural hazards in Indonesia: A view from space and from the households of Nusa Lembongan Island. *Ocean and Coastal Management*, *207*, 105586. https://doi.org/10.1016/j.ocecoaman.2021.105586

Bajpai, K., Student, M., & Soni, R. (2017). Analysis of image enhancement techniques used in remote sensing satellite imagery. *International Journal of Computer Applications, 169*(10), 1-11. http://dx.doi.org/10.5120/ijca2017914884

BPS Provinsi Sulawesi Selatan. (2021). *Hasil Survei Komoditas Perikanan Potensi Rumput Laut Provinsi Sulawesi Selatan 2021*. Badan Pusat Statistik (BPS).

Cheng, J., Jia, N., Chen, R., Guo, X., Ge, J., & Zhou, F. (2022). High-resolution mapping of seaweed aquaculture along the Jiangsu Coast of China using google earth engine (2016-2022). *Remote Sensing*, *14*(24), 6202. https://doi.org/10.3390/rs14246202

Colliot, O. (2023). *Machine learning for brain disorders*. Humana Press. https://doi.org/10.1007/978-1-0716-3195-9

Dang, L., Weng, L., Dong, W., Li, S., & Hou, Y. (2022). Spectral-spatial attention transformer with dense connection for hyperspectral image classification. *Computational Intelligence and Neuroscience*, *2022*, 7071485. https://doi.org/10.1155/2022/7071485

Dora, J. T., Lindberg, S. K., James, P., & Wang, X. (2024). Assessing the potential of fluorescence as a monitoring tool for reproductive tissue in selected macroalgae species. *Journal of Applied Phycology, 36*, 2153-2159. https://doi.org/10.1007/s10811-024-03211-3

Jin, R., Ye, Z., Chen, S., Gu, J., He, J., Huang, L., Christakos, G., Agusti, S., Duarte, C. M., & Wu, J. (2023). Accurate mapping of seaweed farms with high-resolution imagery in China. *Geocarto International*, *38*(1), 2203114. https://doi.org/10.1080/10106049.2023.2203114

Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, *65*, 101759. https://doi.org/10.1016/J.MEDIA.2020.101759

Kington, J., & Collison, A. (2022). *Scene level normalization and harmonization of planet dove imagery*. San Francisco, USA: Planet Labs.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). *Segment anything*. https://doi.org/10.48550/arXiv.2304.026433

Langford, A., Waldron, S., Sulfahri, & Saleh, H. (2021). Monitoring the COVID-19-affected Indonesian seaweed industry using remote sensing data. *Marine Policy*, *127*, 104431. https://doi.org/10.1016/j.marpol.2021.104431

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

Li, C., Liu, Y., Yin, H., Li, Y., Guo, Q., Zhang, L., & Du, P. (2021). *Attention residual U-Net for building segmentation in aerial images* (pp. 4047-4050). In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium. https://doi.org/10.1109/IGARSS47720.2021.9554058

Liu, J., Lu, Y., Guo, X., & Ke, W. (2023). A deep learning method for offshore raft aquaculture extraction based on medium resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16*, 6296-6309. https://doi.org/10.1109/JSTARS.2023.3291499

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing, 152*, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Marquez, L., Fragkopoulou, E., Cavanaugh, K. C., Houskeeper, H. F., & Assis, J. (2022). Artificial intelligence convolutional neural networks map giant kelp forests from satellite imagery. *Scientific Reports*, *12*(1), 22196. https://doi.org/10.1038/s41598-022-26439-w

Mei, X., Pan, E., Ma, Y., Dai, X., Huang, J., Fan, F., Du, Q., Zheng, H., & Ma, J. (2019). Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing*, *11*(8), 963. https://doi.org/10.3390/rs11080963

Nurdin, N., Alevizos, E., Syamsuddin, R., Asis, H., Zainuddin, E. N., Aris, A., Oiry, S., Brunier, G., Komatsu, T., & Barillé, L. (2023). Precision aquaculture drone mapping of the spatial distribution of Kappaphycus alvarezii biomass and carrageenan. *Remote Sensing*, *15*(14), 3674. https://doi.org/10.3390/rs15143674

Osco, L. P., Wu, Q., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). *The segment anything model (SAM) for remote sensing applications: From zero to one shot*. http://arxiv.org/abs/2306.16623

Planet. (2023). *Planet imagery product specifications*. San Francisco, USA: Planet Labs.

Pratama, I., & Albasri, H. (2021). Mapping and estimating harvest potential of seaweed culture using Worldview-2 Satellite images: A case study in Nusa Lembongan, Bali - Indonesia. *Aquatic Living Resources*, *34*, 15. https://doi.org/10.1051/alr/2021015

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, *9351*, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Takahashi, R., Matsubara, T., & Uehara, K. (2015). Data augmentation using random image cropping and patching for deep CNNs. *Journal Of Latex Class Files*, *14*(8), 1-16. https://doi.org/10.1109/TCSVT.2019.2935128

Wang, X., Jing, S., Dai, H., & Shi, A. (2023). High-resolution remote sensing images semantic segmentation using improved UNet and SegNet. *Computers and Electrical Engineering*, *108*, 108734. https://doi.org/10.1016/j.compeleceng.2023.108734

Wang, X., Wang, X., Zhao, K., Zhao, X., & Song, C. (2022). FSL-Unet: Full-scale linked Unet with spatial-spectral joint perceptual attention for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 5539114. https://doi.org/10.1109/TGRS.2022.3208125

Wu, Q., & Osco, L. P. (2023). Samgeo: A Python package for segmenting geospatial data with the segment anything model (SAM). *Journal of Open Source Software*, *8*(89), 5663. https://doi.org/10.21105/joss.05663

Yan, S., Xu, L., Yu, G., Yang, L., Yun, W., Zhu, D., Ye, S., & Yao, X. (2021). Glacier classification from Sentinel-2 imagery using spatial-spectral attention convolutional model. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102445. https://doi.org/10.1016/j.jag.2021.102445

Zhao, Z., Chen, Y., Li, K., Ji, W., & Sun, H. (2024). Extracting photovoltaic panels from heterogeneous remote sensing images with spatial and spectral differences. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 5553-5564. https://doi.org/10.1109/JSTARS.2024.3369660

Zhu, M., Jiao, L., Liu, F., Yang, S., & Wang, J. (2021). Residual spectral-spatial attention network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(1), 449-462. https://doi.org/10.1109/TGRS.2020.2994057