# Deeplabv3+ for extracting Enteromorpha prolifera from drone images

Yun Peng
Shandong University of Science and Technology, Shandong 266590, China
y18366180850@163.com

**Abstract.** The coastal waters of Qingdao have been invaded by Enteromorpha proliferata for 15 consecutive years, causing enormous economic losses and becoming one of the hot spots in marine ecology research in China in recent years. At present, the method used for extracting Enteromorpha prolifera commercially is still manual labeling, which consumes labor and time. In addition, although in-depth learning has developed rapidly in the field of image semantics segmentation, the irregularity of Enteromorpha proliferata and the highly affine transformation brought by the UAV shooting perspective will reduce the accuracy and robustness of in-depth learning methods. To solve this problem, this paper uses Deeplabv3+ semantics segmentation model to extract Enteromorpha prolifera from unmanned aerial vehicle images. Void convolution in Deeplabv3+ can improve the recognition ability of Irregular Enteromorpha prolifera, and spatial pyramid architecture can improve the robustness of the model. Experiments show that the recognition results of the algorithm used in this paper coincide well with those of visual interpretation, and can distinguish between Enteromorpha prolifera and floating objects on the sea such as ships. Deeplabv3+ improves the efficiency of extracting Enteromorpha proliferata and is real-time. Combining with unmanned aerial imagery can reduce the cost of harnessing marine ecological problems.

**Keywords:** deeplabv3+; extraction of enteromorpha prolifera; irregular object segmentation.

## 1. Introduction

Enteromorpha prolifera is a large plankton algae in the ocean. Enteromorpha prolifera inhibits the absorption of dissolved oxygen by other marine organisms and destroys the marine ecological balance. Enteromorpha prolifera inhibits the absorption of dissolved oxygen by other marine organisms and destroys the marine ecological balance. It produces a bad odor when it loses water, which affects the coastal environment and the development of tourism. Widespread coverage of the sea affects fishery development. The coastal area of Qingdao has been threatened by Enteromorpha proliferata for 15 consecutive years. The ecological and economic problems caused by Enteromorpha proliferata should not be underestimated. Therefore, real-time high-precision extraction of Enteromorpha prolifera is important for intelligent decision-making and fast and accurate salvage.

For the existing methods in general, the monitoring of Enteromorpha proliferata is mainly based on the remote sensing images of monitoring vessels and satellites[1,2]. The monitoring vessel can accurately determine the location and scale of Enteromorpha proliferata, but the distribution of Enteromorpha proliferata is widespread and dispersed. It is difficult to know the distribution of Enteromorpha proliferata in real time only through the vessel data, and it is not possible to detect Enteromorpha proliferata in time[3]. In order to automate the extraction of Enteromorpha proliferata, remote sensing images are widely used in this task. Fuxiang Xu et al. used GF1-WFV data to quantitatively study the monitoring error of MODIS data, obtained the error distribution range, and analyzed the adaptability of MODIS data[4]. Cui, B. et al. applied the training results from GF1-WFV to MODIS and proposed a deep semantic semantically segmented network (SRSe-Net), which effectively obtains detailed green tide boundary information using fewer network parameters[5]. X. Wan et al. proposed the EP rough-then-accurate extraction network, which considers sample balance, pixel learning spectral information, and context dependence between pixels. This method of Enteromorpha proliferata identification has strong spatial adaptability[6]. However, Enteromorpha proliferata is a fine-grained irregular object, and the resolution of remote sensing image limits its representation. In addition, continuous and advanced in-depth learning methods are rarely used in the field of Enteromorpha proliferata extraction.

Due to the limited resolution of remote sensing images, this article considers drone imaging. At present, infrared imaging patrol system is formed on the unmanned aerial vehicle, which can effectively avoid the influence of visible light, environment and weather. Single-mode and multi-mode fusion patrol of target Enteromorpha prolifera independently or in combination with visible light can obtain high-resolution image of Enteromorpha prolifera[7]. By preprocessing the original image from the unmanned aerial vehicle, a semantically segmented dataset of Enteromorpha prolifera based on the unmanned aerial vehicle image is constructed, and the deep learning model is trained and tested on the corresponding dataset.

In the field of computer vision, in-depth learning continues to develop, and is superior to traditional image processing methods in accuracy and efficiency. It has been widely used in target detection and image segmentation tasks. In recent years, Semantic Segmentation Model mainly focuses on the improvement of the accuracy, multi-scale and segmentation efficiency of classical codec-decode models, such as FCN[8], U-Net[9], SegNet[10], PSPNet[11], DeepLab series[12-14], etc. Among them, DeepLab series is the most used algorithm in the field of semantic segmentation. DeepLabv1 introduces void convolution into the semantic segmentation network to increase the perception field of the network by increasing the void rate of the convolution core[15]. On the basis of DeepLabv1, DeepLabv2 uses the network design idea of SPPNet[16] for reference, and proposes a spatial pyramid down sampling structure. Based on DeepLabv2, DeepLabv3 improves the sampling structure under the spatial pyramid. DeepLabv3+ is the best performance algorithm in the DeepLab family. It is different from the common symmetric structure of encoder and decoder. DeepLabv3+ uses Xception[17] structure as the network backbone. The main parameters of the network structure are concentrated in the encoder part, while the decoder only has two quadruple sampling and interpolation operations, making it more suitable for high resolution image segmentation and effectively identifies the fine-grained irregular object Enteromorpha prolifera.

For unmanned aerial vehicle images, the irregularity of Enteromorpha proliferata and the highly affine transformation brought by the view angle of the unmanned aerial vehicle will reduce the accuracy and robustness of most deep learning semantics segmentation methods. Considering that the hollow convolution in Deeplabv3+ can improve the recognition ability of irregular Enteromorpha prolifera, the spatial pyramid architecture can improve the robustness of the model. Therefore, this paper takes the extraction of Enteromorpha prolifera from the coastal waters of Qingdao as a semantic segmentation task, chooses the unmanned aerial vehicle that can obtain high spatial resolution image as an experimental platform, and uses Deeplabv3+ model to achieve efficient and accurate extraction of Enteromorpha prolifera, thereby reducing the cost of harnessing marine ecological problems.

## 2. Basic theory

### 2.1 Deeplabv3+ Model

The Deeplabv3+model introduces an encoder and decoder structure, which combines ASPP structure with Encoder-Decoder method [18]. It can be used not only to obtain multi-scale image context information, but also to capture image boundaries by reconstructing spatial information. For ASPP's output feature map $x$, its output feature map $(y)$ is:

$$y = \sum_{m-1}^{M} x[a + rm]\,\omega[m] \tag{1}$$

Formula: $a$ is a one-dimensional input signal; $\omega[m]$ is the filter value; $r$ is the expansion rate; $M$ is the length of the void convolution.

DeepLabv3+ improves image recognition compared to previous DeeppLabv3, which is directly related to the decoding module added to it and improves the image boundary processing.

DeepLabv3+ adds the void convolution method, which expands on the original convolution. With exponential expansion, the multi-scale image context information can be aggregated effectively without affecting the image resolution. Integrating all previous semantics segmentation algorithms will improve the accuracy of image extraction system and achieve the desired results. The overall model architecture is shown in Figure 1.

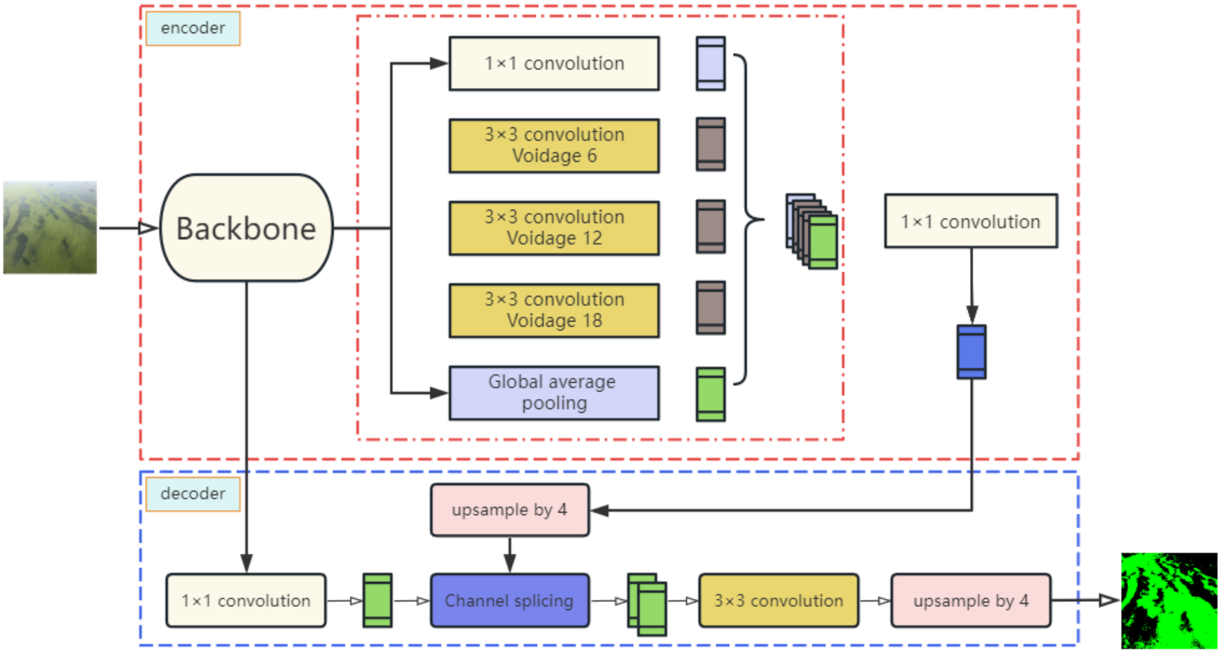



**Figure 1.** Deeplabv3+ Diagram

## 3. Experimental method

Although in-depth learning has been widely used in image semantics segmentation, due to the diversity and complexity of the marine environment, and the irregularity of Enteromorpha prolifera itself, there is still a problem of low accuracy in general image segmentation methods for Unmanned Aerial Enteromorpha prolifera images. Therefore, we built the extraction pipeline of Enteromorpha proliferata based on Deeplabv3+. We first constructed the Enteromorpha prolifera data set and trained Deeplabv3+. In the testing phase, we process the image to get the final test recognition result by using the enhanced image as input to Deeplabv3+.

### 3.1 Construction of Segmented Pipeline Based on Deeplabv3+

We built a pipeline based on Deeplabv3+for extracting Enteromorpha prolifera from drone images, which is summarized as follows: 1. Data set making: Select representative image samples of Enteromorpha proliferata, preprocess them, and label them. 2. Model building: build Deeplabv3+, use the dataset completed by the above steps, randomly divide the training set according to 7:3 for model training. After meeting the required test precision, terminate the training. 3. In the test phase, the original image is processed, and the enhanced image is used as the input of Deeplabv3+, and the final test recognition result is obtained. The pipeline flowchart presented in this paper is shown in Figure 2.

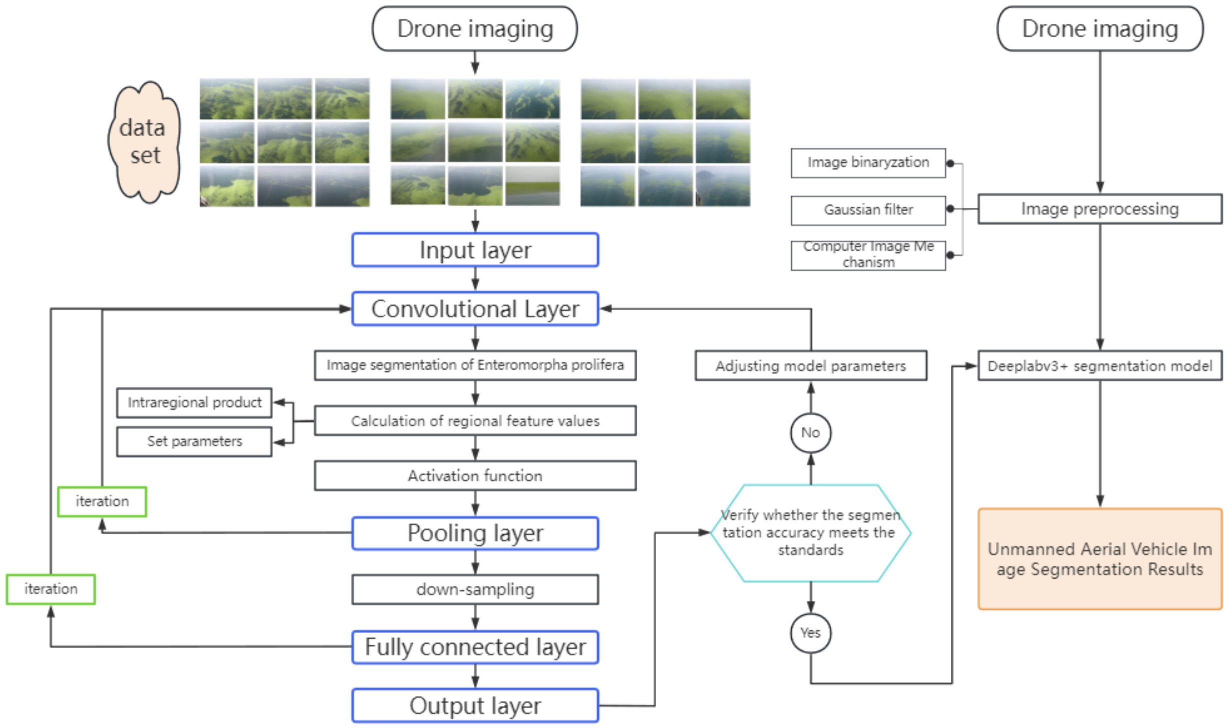



**Figure 2.** Drone Enteromorpha Image Recognition Pipeline

### 3.2 Dataset production

Before building a neural network, it is necessary to first create the corresponding dataset.In order to make the Deeplabv3+ model training have good generalization ability and robustness, and improve the accuracy of image segmentation, a large number of related datasets are required for training.

Due to the short rise time of unmanned aerial vehicle monitoring of Enteromorpha prolifera, the number of Enteromorpha prolifera images is insufficient to support model training. Therefore, the ADE20K open source dataset was selected for model training. The ADE20K dataset contains 20410 images, covering 89% of pixels and 150 categories, making it universally applicable.

(1) Divide the dataset into a training set and a testing set, and set them up. Extract 90% of the existing unmanned aerial vehicle Enteromorpha video images as the training set, with the rest being the testing set;

(2) Use the commonly used labelme label making tool for deep learning semantic segmentation to create labels for the training and testing sets respectively;

(3) Name the label and the original image set according to unified rules, and prepare for training and testing.

### 3.3 Building Tensorfolw based on Deeplabv3+

With the development of deep learning, many methods for building convolutional neural networks have emerged, such as Pytorch, TensorFlow, MXNet, etc. Different building frameworks have different advantages. This experiment adopts TensorFlow to build a convolutional neural network. The TensorFlow workflow is easy to understand, has strong compatibility, is flexible and efficient, and is completely free of charge with open elements. It has been widely used in the field of image segmentation.

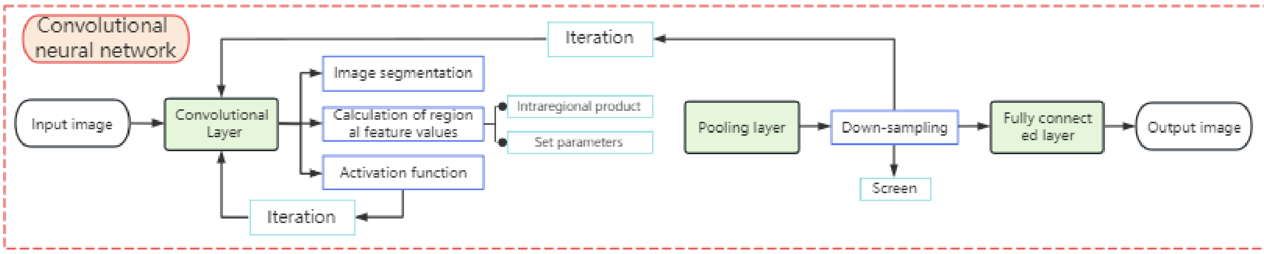



**Figure 3.** Schematic diagram of Tensorfolw network construction

Prediction section:

(1) Using the Xception series as the backbone feature extraction network, dimensionality is increased using $1\times1$ convolution, followed by feature extraction using $3\times3$ deep separable convolutions, and finally dimensionality is reduced using $1\times1$ convolutions [19].

(2) After completing feature extraction, strengthen the feature extraction structure. In the encoder, the initially effective holes in the feature layer are convolved, the Receptive field is expanded, feature extraction and merging are performed, and the $1\times1$ convolution feature compression is completed.

(3) Use the features obtained above to obtain prediction results.

Training section:

(1) Training file parsing, training the network by comparing the probability of each pixel category in the predicted results with labels.

(2) Loss parsing.

### 3.4 Image enhancement

(1) Using image binarization for first image enhancement:

For an image, assume that the proportion of foreground pixels to the total number of pixels in the image is $\omega_0$, with a mean of $u_0$; Set the proportion of background pixels to the total number of pixels in the image to $\omega_1$, with an average of $u_1$; The average grayscale of the modified image is

$$u=\omega_0\times u_0+\omega_1\times u_1 \tag{2}$$

According to the above formula, convert the RGB image of Enteromorpha prolifera into a grayscale image to enhance the color contrast between Enteromorpha prolifera and seawater.

Based on the maximum inter class variance objective function of grayscale images

$$g(t)=\omega_0\times(u_0-u)^2+\omega_1\times(u_1-u)^2 \tag{3}$$

The larger the $g(t)$, the greater the probability of separating the image background from the target, and the better the separation effect. However, according to the images we processed, due to lighting factors, some of the Enteromorpha prolifera and seawater have similar colors, and the boundary division is not clear. Therefore, using image binarization alone for image enhancement is not effective.

(2) Use Gaussian filter to enhance the image for the second time:

Gaussian filter is a method of weighted averaging the whole image. The value of each pixel is obtained by weighted average of its own and other pixel values in the neighborhood[20].

Gaussian filter is used to enhance the image, and it is found that the image has improved in detail, and the clutter is significantly reduced, but the overall color contrast has not improved significantly, and the overall boundary segmentation is not obvious, so the image enhancement effect of Gaussian filter alone is not good.

(3) Using computer image mechanism for the third image enhancement:

In computer image recognition, it includes information acquisition, preprocessing, feature extraction and selection, classifier design and classification decision-making. The computer image is composed of RGB three primary colors, and RGB threshold analysis is performed on the drone Enteromorpha prolifera image, it was found that the B value of the seawater image is greater than the

G value, while the G value of the Enteromorpha prolifera image is greater than the B value, and the G value of the Enteromorpha prolifera image is higher than 100 but the B value is lower than 100. This is used as a color threshold limit to traverse the image pixels, making the seawater part turn into white pixels and retaining the Enteromorpha prolifera image. After processing, it was found that the edge segmentation of the Enteromorpha prolifera image is good, but there are many internal clutter points in the Enteromorpha prolifera image, and the segmentation is relatively chaotic. Therefore, using computer image mechanism alone for image enhancement is not effective.

(4) Combining the first three methods for the fourth image enhancement:

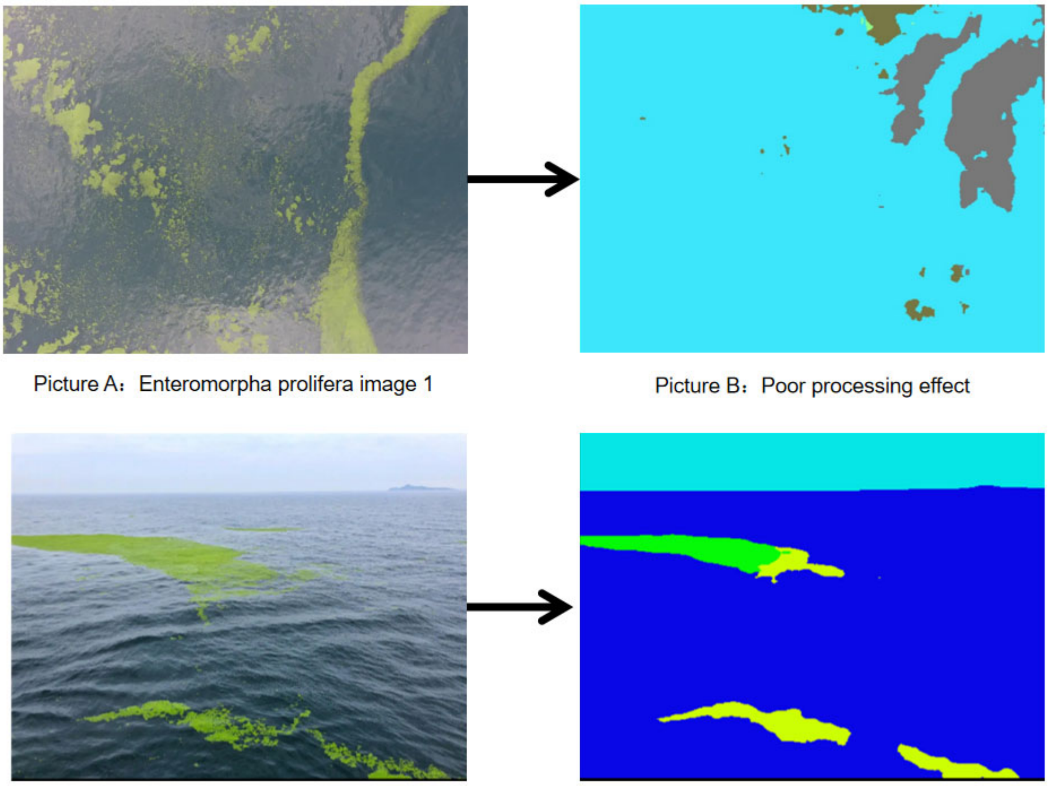
After the first three image augmentations, it was found that each image enhancement method has its own advantages and disadvantages. We will combine the first three image enhancement methods. First, enhance the color contrast of the image, and filter the noise with Gaussian filter to make the whole image more smooth. Finally, use the computer image correlation mechanism to increase the condition of color threshold limit, traverse the image pixels, desalinate the seawater area, and make it become white pixels. After the above image enhancement processing, the image is substituted into a neural network to obtain the final segmentation image of Enteromorpha prolifera.

After discussion, it was found that the combination of multiple image enhancement methods significantly improved the image segmentation effect.

## 4. experimental result

### 4.1 Image segmentation of Enteromorpha prolifera

By substituting drone images of Enteromorpha prolifera into the trained Deeplabv3+ model, the analysis results showed that the extraction effect was poor for images with low resolution and dispersed distribution of Enteromorpha prolifera (as shown in Figures A and B), and good for images with high resolution and small area of Enteromorpha prolifera (as shown in Figures C and D).



Picture A：Enteromorpha prolifera image 1 Picture B：Poor processing effect

Picture C：Enteromorpha prolifera image 2 Picture D：Good processing effect

**Figure 4.** Comparison before and after image segmentation

Based on the above situation, image enhancement is added throughout the entire extraction process to enhance the color contrast of the image, highlight the Enteromorpha section, and reduce the difficulty of segmentation.

Preprocess the unmanned aerial vehicle's Enteromorpha prolifera image with color enhancement, pixel traversal, etc., and substitute it into the trained Deeplabv3+ model.

Firstly, the input drone Enteromorpha prolifera image data is uniformly sized and cropped, and the cropped image is subjected to image enhancement operations such as random color, noise addition, and pixel traversal (As shown in Figures b and c); Secondly, the improved convolutional neural network was fed for feature extraction and multiple iterations. At this stage, the image data information was subjected to multi-scale feature extraction and channel weighting processing, obtaining good global and local features; Finally, it is substituted into the trained Deeplabv3+ model, and the obtained features are upsampled to recover to the original image size and classified pixel by pixel to achieve the goal of semantic segmentation of the target (As shown in Figure d).

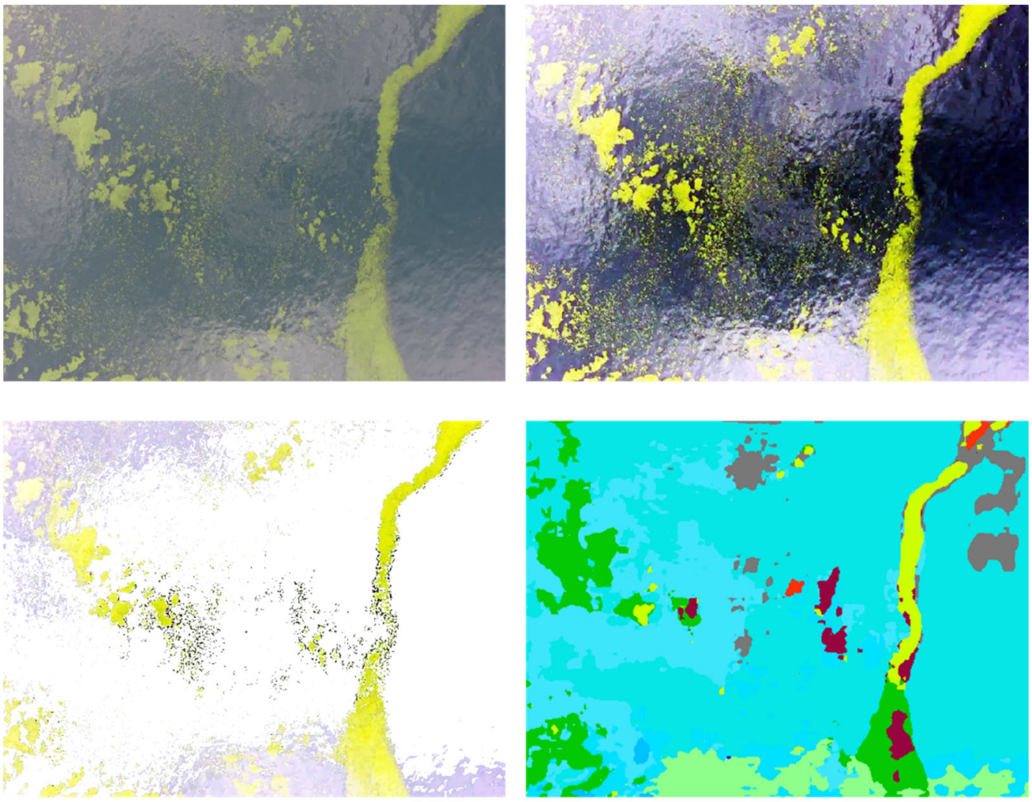


**Figure 5.** Schematic diagram of image enhancement process and results

Moreover, the model can accurately extract Enteromorpha prolifera even with external influencing factors. As shown in Figure Examples 1 and 2, only seawater and Enteromorpha prolifera are distributed in the figure. Under the semantic segmentation algorithm model of Deeplabv3+, the Enteromorpha prolifera segmentation effect is obvious; As shown in Figure Example 3, there is a ship at the bottom of the figure, and the model can still distinguish between Enteromorpha prolifera and other ground objects except for seawater, accurately identifying Enteromorpha prolifera, and achieving segmentation.

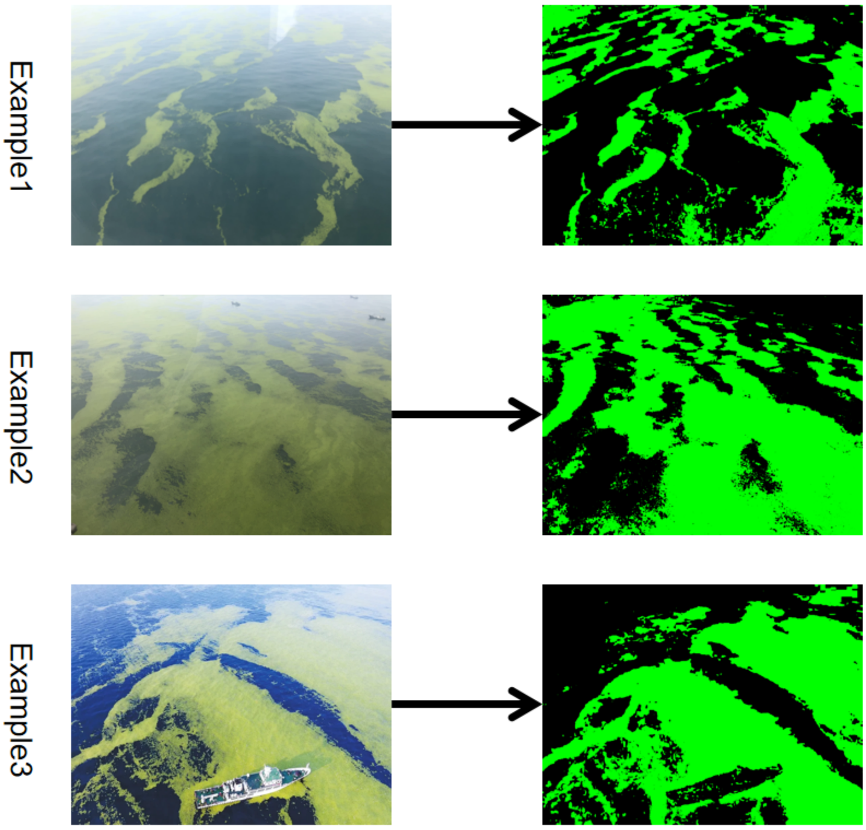



**Figure 6.** Visualization results on Deeplabv3+model validation set

### 4.2 Discuss

Compared with manual visual interpretation, the semantic segmentation algorithm model based on Deeplabv3+ can achieve a coincidence degree of over 85%. For floating objects on the sea such as ships, it can be well segmented with high overall accuracy. However, when processing nearshore images, the vegetation coverage along the coast can affect model recognition.

Compared with traditional manual visual interpretation, the semantic segmentation algorithm model based on Deeplabv3+ achieves automatic segmentation of Enteromorpha prolifera images by computers, greatly improving the speed of Enteromorpha prolifera extraction, improving the overall efficiency of business operation, reducing labor investment, achieving real-time Enteromorpha prolifera warning, and reducing economic losses.

The contribution of this article to the task of dividing Enteromorpha prolifera is: Using drones as an observation platform, large-scale and refined observation of Enteromorpha prolifera is carried out, providing a large amount of high-quality data for Enteromorpha prolifera recognition. Add image preprocessing stage, use various image enhancement methods such as random color, noise, and pixel traversal to increase the feature difference between Enteromorpha prolifera and seawater, and enhance the interpretation and recognition effect of Enteromorpha prolifera. Adding Decoder technology to DeepLabv3+ further integrates the low-level and high-level features of Enteromorpha prolifera images, further improving the accuracy of edge segmentation and making the edge features of Enteromorpha prolifera and other objects clearer.

## 5. Conclusion

This article proposes a semantic segmentation algorithm model based on Deeplabv3+ for the problem of unmanned image segmentation of Enteromorpha prolifera.Using drone photography to obtain images of Enteromorpha prolifera, and image binarization, Gaussian filter, image threshold analysis were used to enhance the image of Enteromorpha prolifera to improve the image contrast;

Substitute drone images of Enteromorpha prolifera into the trained Deeplabv3+ model to achieve automatic recognition of Enteromorpha prolifera images. This study aims to solve the problems of slow manual visual interpretation speed, low efficiency, and large workload, while reducing labor costs while also taking into account high-precision recognition of Enteromorpha prolifera images. This provides an effective research method for efficient, low-cost, and large-scale image extraction of Enteromorpha prolifera in Qingdao, Shandong, Lianyungang, Jiangsu, and other places. However, there are still some issues that need further research: (1) Plants with similar spectral characteristics along the coast are easily misidentified; (2) Further optimize the parameters of the DeepLabv3+ model; (3) Further research is needed to determine whether different models of unmanned aerial vehicles have universality; (4) Further research is needed to improve the extraction accuracy of Enteromorpha prolifera.

## References

[1] Zhou M-J, Liu D-Y, Anderson D M, et al. Introduction to the Special Issue on green tides in the Yellow Sea[J]. Estuarine, Coastal and Shelf Science, 2015, 163: 3-8.

[2] Valiela I, Mcclelland J, Hauxwell J, et al. Macroalgal blooms in shallow estuaries: controls and ecophysiological and ecosystem consequences[J]. Limnology and oceanography, 1997, 42(5part2): 1105-1118.

[3] Fletcher R. The occurrence of “green tides”—a review[J]. Marine benthic vegetation: recent changes and the effects of eutrophication, 1996: 7-43.

[4] Xu F, Gao Z, Ning J, et al. Error analysis on Green Tide monitoring using MODIS data in the Yellow Sea based on GF-1 WFV data[C]. Remote Sensing and Modeling of Ecosystems for Sustainability XIII, 2016: 191-196.

[5] Cui B, Zhang H, Jing W, et al. SRSe-net: Super-resolution-based semantic segmentation network for green tide extraction[J]. Remote Sensing, 2022, 14(3): 710.

[6] Wan X, Wan J, Xu M, et al. Enteromorpha coverage information extraction by 1D-CNN and Bi-LSTM networks considering sample balance from GOCI images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 9306-9317.

[7] Lelong C C, Burger P, Jubelin G, et al. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots[J]. Sensors, 2008, 8(5): 3557-3585.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 3431-3440.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015: 234-241.

[10] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[11] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 2881-2890.

[12] Chen L-C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.

[13] Chen L-C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[14] Chen L-C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[15] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.

[16] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[17] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1251-1258.

[18] Chen L-C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 801-818.

[19] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional networks[C]. Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS): 1106-1114.

[20] Li K, Zhang Y-S, Tong X-C, et al. An edge detection algorithm based on anisotropy and edge strength correction factor[J]. Computer Engineering & Science, 43(07): 1256.