

Mangrove monitoring and extraction based on multi-source remote sensing data: a deep learning method based on SAR and optical image fusion

Yiheng Xie¹, Xiaoping Rui^{1*}, Yarong Zou^{2,3*}, Heng Tang¹, Ninglei Ouyang¹

¹ School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

² National Satellite Ocean Application Center, Ministry of Natural Resources, Beijing 100081, China

³ Key Laboratory of Space and Ocean Remote Sensing and Application Research, Ministry of Natural Resources, Beijing 100081, China

Received 9 January 2024; accepted 24 June 2024

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Mangroves are indispensable to coastlines, maintaining biodiversity, and mitigating climate change. Therefore, improving the accuracy of mangrove information identification is crucial for their ecological protection. Aiming at the limited morphological information of synthetic aperture radar (SAR) images, which is greatly interfered by noise, and the susceptibility of optical images to weather and lighting conditions, this paper proposes a pixel-level weighted fusion method for SAR and optical images. Image fusion enhanced the target features and made mangrove monitoring more comprehensive and accurate. To address the problem of high similarity between mangrove forests and other forests, this paper is based on the U-Net convolutional neural network, and an attention mechanism is added in the feature extraction stage to make the model pay more attention to the mangrove vegetation area in the image. In order to accelerate the convergence and normalize the input, batch normalization (BN) layer and Dropout layer are added after each convolutional layer. Since mangroves are a minority class in the image, an improved cross-entropy loss function is introduced in this paper to improve the model's ability to recognize mangroves. The AttU-Net model for mangrove recognition in high similarity environments is thus constructed based on the fused images. Through comparison experiments, the overall accuracy of the improved U-Net model trained from the fused images to recognize the predicted regions is significantly improved. Based on the fused images, the recognition results of the AttU-Net model proposed in this paper are compared with its benchmark model, U-Net, and the Dense-Net, Res-Net, and Seg-Net methods. The AttU-Net model captured mangroves' complex structures and textural features in images more effectively. The average OA, F1-score, and Kappa coefficient in the four tested regions were 94.406%, 90.006%, and 84.045%, which were significantly higher than several other methods. This method can provide some technical support for the monitoring and protection of mangrove ecosystems.

Key words: image fusion, SAR image, optical image, mangrove, deep learning, attention mechanism

Citation: Xie Yiheng, Rui Xiaoping, Zou Yarong, Tang Heng, Ouyang Ninglei. 2024. Mangrove monitoring and extraction based on multi-source remote sensing data: a deep learning method based on SAR and optical image fusion. *Acta Oceanologica Sinica*, 43(9): 110–121, doi: 10.1007/s13131-024-2356-1

1 Introduction

Mangroves are a special vegetation that grows in intertidal zones and are usually distributed in coastal wetlands (Twilley, 2019). They have important functions such as salt tolerance, storm surge resistance, habitat provision, and protection of coastal ecosystems. The mangrove ecosystem is one of the richest biodiversity systems on Earth, providing rich fishery resources for coastal areas, maintaining water quality and soil stability, and regulating global climate change (Wang et al., 2021b). Therefore, more accurate and rapid extraction of mangrove vegetation information from images is significant for mangrove monitoring and protection (Maurya et al., 2021; Giri, 2016).

Traditional methods for mangrove vegetation extraction include artificial visual interpretation, index methods, and image classification based on texture and shape features (Darko et al.,

2021; Maurya et al., 2021). First, although the manual visual interpretation method is intuitive and easy to understand, experienced interpreters can achieve high interpretation accuracy with professional knowledge and experience (Mahmoud, 2012; Braun, 2021). However, this method is time-consuming, expensive, and requires considerable human resources. Moreover, because of the particularity of the mangrove growing environment, it is difficult for conventional field investigations to meet the monitoring requirements of mangroves with high spatial and temporal resolutions (Zhang et al., 2021; Lu and Wang, 2021). Second, although the Index method is simple and easy to implement, for example, by calculating different indices of remote sensing images (such as normalized difference vegetation index, NDVI) (Huang et al., 2021), a preliminary classification of mangrove vegetation areas can be realized. Mangrove areas have complex vegetation struc-

Foundation item: The Key R&D Project of Hainan Province under contract No. ZDYF2023SHFZ097; the National Natural Science Foundation of China under contract No. 42376180.

*Corresponding author, E-mail: ruixp@hhu.edu.cn; zyr@mail.nsoas.org.cn

tures and surface features (Kamal et al., 2014; Cao et al., 2018), including canopies, water bodies, and mud bogs. It is difficult for a classification method based on the index method to distinguish these complex features effectively because it relies primarily on a simple combination of spectral information (Tran et al., 2022; Maurya et al., 2021).

Firstly, image classification methods based on texture and shape features include machine and deep-learning methods (Gonzalez-Perez et al., 2022). In mangrove vegetation identification research, machine learning methods include the support vector machine (SVM), random forest (RF), and K-nearest neighbors (KNNs) (Sandra and Rajitha, 2023; Cao et al., 2018). First, the SVM principle is simple, requires no parameter adjustments, and has a good generalization ability and recognition accuracy for small-scale datasets (Wang et al., 2021a). However, SVM is based on hand-extracted features for classification, limiting learning of complex abstract features (Fu et al., 2023; Luo et al., 2017). Simultaneously, the SVM method often fails to meet research expectations for modeling nonlinear relationships (Toosi et al., 2019; Raghavendra and Deka, 2014).

Second, the random forest has little influence on outliers and can reduce overfitting to a certain extent because it is an ensemble learning method based on multiple decision trees (Xu et al., 2023b; Shen et al., 2023). However, random forests learn features at shallow levels, and it is difficult for them to learn higher-level and more abstract feature representations automatically. It has the same problem as the SVM method because the random forest is an integrated method based on a decision tree; therefore, the modeling of complex nonlinear relations in images is limited.

Finally, KNN is an intuitive and easy-to-understand algorithm without a complex model structure and parameter adjustment and can effectively reduce the model construction time (Su et al., 2023; Tian et al., 2023). However, the calculation cost is higher than the SVM and RF methods. When making predictions, the KNN must calculate the distance between the test samples and all the training samples. Moreover, KNN is very sensitive to outliers because its prediction results are affected by the nearest neighbor samples, and a single outlier may have a greater impact on the prediction results. In summary, the mangrove image classification method based on machine learning can be applied better in complex environments than index methods, and has high spatiotemporal resolution monitoring and big data processing capabilities (Maurya et al., 2021). However, this requires manual feature engineering and is limited to shallow feature learning.

More studies have introduced deep learning techniques in mangrove vegetation identification to overcome these limitations and improve the accuracy and automation of mangrove identification (Xu et al., 2023a; Wei et al., 2023). At the method level, these studies used convolutional neural networks (CNNs) to extract deep features accurately. As this method is not limited by the size of the input image, it exhibits strong robustness and portability; therefore, it is very popular in image semantic segmentation. The improved U-Net network was used to classify mangrove vegetation based on GF-2 cloudless and unobserved optical image, and the average overall accuracy reached 94.43% (Yu et al., 2023). The precision of the literature can reach 92.0%, but the cloudless and unobserved optical image is also selected (Wei et al., 2023). Therefore, the main research data source at the image level is high-resolution remote-sensing images, relying mainly on optical information. Although optical images have high spatial resolution and rich color information, they are easily

limited by weather and lighting conditions (Yang et al., 2022). Simultaneously, optical images cannot penetrate clouds, vegetation cover, or underground structures; therefore, the information obtained under complex geomorphological conditions, such as mangrove forests, is incomplete. Synthetic aperture radar (SAR) images can penetrate clouds and collect information under bad weather conditions (Purnamasayangasukasih et al., 2016); however, their image details are relatively poor. Therefore, this paper proposed a pixel-level fusion method for SAR and optical images in this study. Fusion images can retain the high resolution and color information of optical images and use the penetration ability of SAR images to provide more detailed and comprehensive features of ground objects, thus enhancing the texture features and shapes of images and improving the recognition accuracy of mangrove forests (Kulkarni et al., 2020; Li et al., 2023).

Regarding research methods, this paper chose the U-Net as a benchmark. First, mangrove vegetation recognition is vulnerable to data sample limitations, and the U-Net network performs well when learning with a few samples (Wei et al., 2023). Good training performance can be obtained using a few labeled samples. Simultaneously, the unique upsampling structure of the U-Net network enables it to retain more spatial information, which is particularly effective for mangrove image segmentation (de Souza Moreno et al., 2023). In identifying mangrove vegetation, retaining spatial information is important for capturing the details and edges of vegetation (Chen et al., 2023). The upsampling structure of U-Net can accurately extract features while maintaining a certain spatial resolution.

For mangrove vegetation identification tasks, although the U-Net has superior performance (Fu et al., 2022), to further improve model generalization, slow down overfitting, enhance attention to important features, and optimize the loss function, this study introduces the dropout layer, batch normalization (BN) layer, attention mechanism, and improved cross-entropy loss function (CLoss) (Xie et al., 2023). First, because mangrove vegetation occupies a small proportion of the image, mangrove categories are rare in the training data (Jia et al., 2019). This imbalance makes the models focus too much on other categories and ignore mangroves, increasing the overfitting of background information (Xu et al., 2023b). Therefore, this paper introduces a dropout layer to force the model not to over-rely on specific neurons by randomly deactivating some neurons during training, improving the generalization and mitigating the overfitting effects. Second, in the identification of mangrove vegetation, the distribution of vegetation varies with changes in the terrain and environment. Gradient disappearance or explosion can easily occur during the training process, affecting the training stability of the model. Therefore, this paper introduced a BN layer to standardize the input of each layer, alleviate the gradient problem, accelerate convergence, and improve the training efficiency and stability of the model. Third, in mangrove vegetation identification, it is essential to focus on the accuracy of the mangrove areas to improve model performance. Therefore, this paper introduces an attention mechanism to make the network focus more on the mangrove vegetation area and improves the model's performance in the target area. Finally, consistent with the reason for adding the dropout layer, mangrove vegetation is a minority category in the overall image, and the traditional cross-entropy loss function causes the model to overlearn the major categories when dealing with an unbalanced category distribution. Therefore, this paper introduces an improved CLoss to weigh the losses of different categories so that the model identifies mangrove vegetation. By adjusting the weights, the model can deal with each

category in a more balanced manner, and the identification accuracy of mangrove vegetation can be improved.

The AttU-Net model for mangrove vegetation recognition under the condition of fused images is thus constructed. In addition, in order to further improve the accuracy of mangrove vegetation extraction, this paper introduces a sliding overlap splicing method to predict the results, which is mainly used to recognize the problem of splicing traces and insufficient edge information in the image. This will improve the accuracy of the mangrove identification model and provide technical support for mangrove ecological protection and management.

2 Research area

The Hainan Dongzhaigang National Nature Reserve is located northeast of Hainan Island at the junction of Haikou and Wenchang cities. Its geographical coordinates are 110°32'–110°37'E and 19°51'–20°10'N. It is a natural reserve of wetland types. Dongzhaigang protected area has a tropical monsoon climate, with an average annual temperature of 23.8°C (28.4°C in July, 17.1°C in January), annual rainfall of 1 700 mm, and frequent typhoons in the rainy season, causing strong winds and torrential rain. The highest sea water temperature is 32.6°C, the lowest is 14.6°C, and the average is 24.5°C.

The Dongzhaigang mangrove reserve has many trees, a large mangrove area, and a favorable ecological environment. The diversity of mangrove forests in the region provides a broader sample for research and helps verify the model's applicability to different mangrove vegetation. Simultaneously, the large distribution of mangroves in the region provides sufficient space and data for a more comprehensive understanding of their structure, function, and dynamic changes. Figure 1 shows a geographical

location map of the study area.

3 Data sources and data pre-processing

3.1 Data sources

In this study, Gaofen-3 (GF-3) satellite SAR image data and Gaofen-6 (GF-6) satellite optical image data from Hainan Island were used to extract high-precision mangrove vegetation using the Mangrove Nature Reserve at the junction of Haikou City and Wenchang City. The GF-3 satellite is a remote-sensing satellite from China's GF-3 Special Project. It is a 1-m resolution radar remote sensing satellite, and it is also China's first C-band multi-polarization SAR imaging satellite with a resolution of 1 m. The GF-3 satellite has 12 imaging modes, including the traditional strip and scanning imaging modes, the wave imaging mode for marine applications, and the global observation imaging mode, the world's most common imaging mode of synthetic aperture radar satellites. Table 1 lists the full-polarization imaging modes and capabilities of the GF-3 SAR images. GF-6 is a low-orbit optical remote-sensing satellite, featuring a combination of high resolution and wide coverage. The GF-6 satellite has a 2-m panchromatic/8-m multispectral high-resolution camera, a 16-m multispectral medium-resolution wide-format camera, a 2-m panchromatic/8-m multispectral camera with an observation width of 90 km, and a 16-m multispectral camera with an observation width of 800 km. Table 2 shows the payloads of the GF-6 satellite.

3.2 Data pre-processing

In this paper, panchromatic and multispectral images of the GF-6 mangrove study area were corrected using radiometric, at-

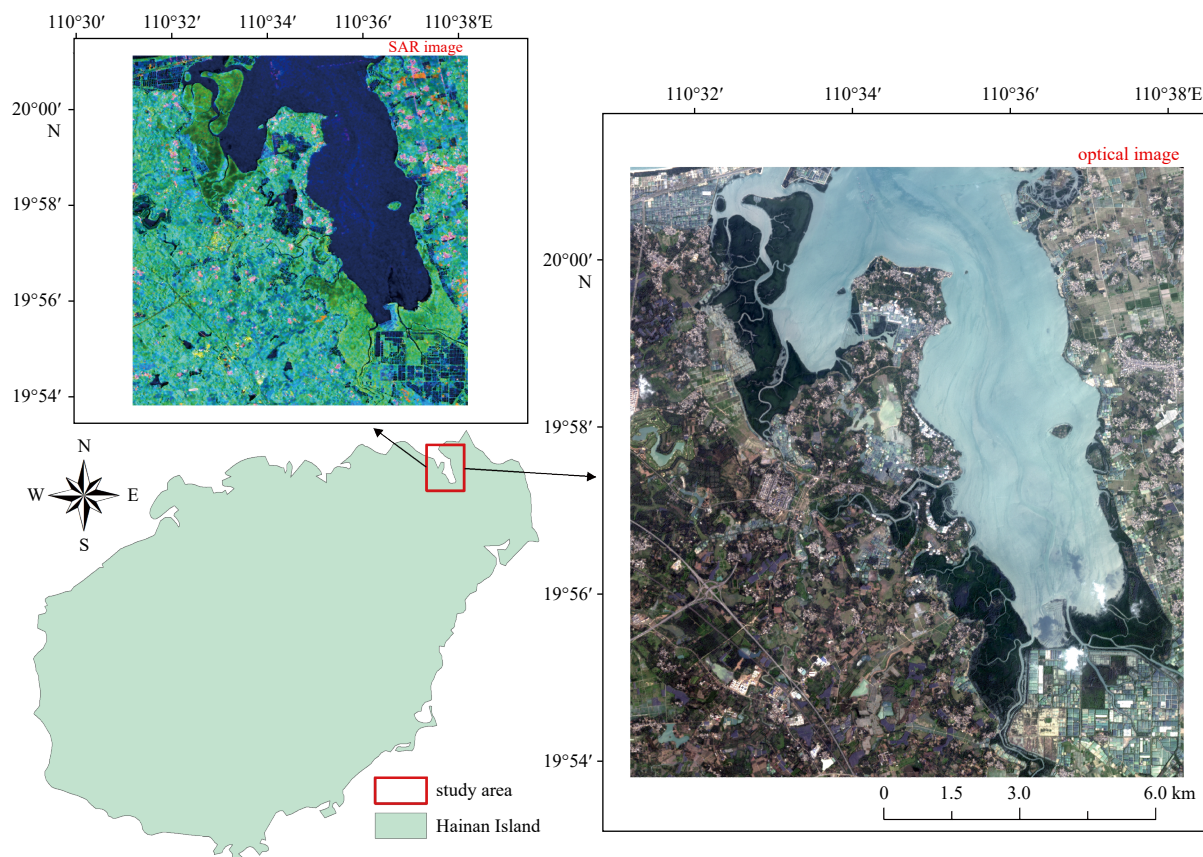


Fig. 1. Geographical location, SAR image, and optical image of the study area.

Table 1. Full polarization imaging mode and capability of Gaofen-3 SAR image

Serial number	Working mode	Angle of incidence/(°)	Visual number $A \times E$	Resolution/m			Imaging bandwidth/km		Polarization mode	Wave position
				Nominal	Azimuth direction	Distance direction	Nominal	Scope		
1	fully polarized band 1	20–41	1 × 1	8	8	6–9	30	20–35	full polarization	Q1–Q28
2	fully polarized band 2	20–38	3 × 2	25	25	15–30	40	35–50	full polarization	WQ1–WQ16
3	wave pattern	20–41	1 × 2	10	10	8–12	5 × 5	5 × 5	full polarization	Q1–Q28

Table 2. Satellite payload of Gaofen-6

Camera type	Band number	Spectrum/ μm	Substellar point pixel resolution	Covering width
Off-axis TMA total reflection type	panchromatic band (P)	0.45–0.90	full color: better than 2 m	>90 km
Off-axis TMA total reflection type	blue spectrum (B1)	0.45–0.52	multispectral: better than 8 m	>90 km
Off-axis TMA total reflection type	green spectrum (B2)	0.52–0.60	multispectral: better than 8 m	>90 km
Off-axis TMA total reflection type	red band (B3)	0.63–0.69	multispectral: better than 8 m	>90 km
Off-axis TMA total reflection type	near-infrared spectrum (B4)	0.76–0.90	multispectral: better than 8 m	>90 km

mospheric and orthometric corrections. Subsequently, the corrected panchromatic and multispectral images were fused to obtain optical images with higher spatial and temporal resolution. The fully polarized SAR incoherent polarization decomposition products are obtained by taking the standard products of L1A class single-view complex images (SLC) of three fully polarized observation modes (fully polarized stripe 1, fully polarized stripe 2, and wave modes) from the GF-3 satellite and the 1-m C-SAR satellite as the inputs, and by the processing steps of Pauli vector transform, polarization coherence matrix transform, fully polarized filtering, and reflection symmetry decomposition. Subsequently, this paper used the optical image of GF-6 as the reference image and the full-polarization SAR target decomposition result for geographic registration. Finally, in this paper, the pre-processed optical images and polarization SAR decomposition results are cropped according to the same region of interest, and the optical and fully polarized decomposed images of the same region are obtained.

In this paper, preprocessed optical and SAR images are fused by code with pixel-level weighting. The fused image can provide richer and more comprehensive surface information. The weights of the SAR and optical images can be set to a and b , re-

spectively, satisfying $a + b = 1$ corresponding to the fusion image in the figure. Here, a represents the weight of the SAR image, and b is the weight of the optical image. By adjusting the proportions of SAR and optical images in the fusion image, this paper divided the study into 11 fusion images with different proportions, as Fig. 2 shows.

The designed weighted fusion module involves two steps. First, due to different pixel sizes, the two images of the same region of interest after cropping will have texture and size mismatch problem. So, to address this problem, the two images are resized. A bilinear interpolation method is used to calculate the values of the new pixels according to the size and pixel layout of the optical image, so that the resized SAR image and the optical image match exactly in size, thus aligning the texture information of the features in the two images. The formula is as follows:

$$\text{dst}(x, y) = (1 - a)(1 - \beta) \cdot \text{src}(c, d) + a(1 - \beta) \cdot \text{src}(c + 1, d) + (1 - a)\beta \cdot \text{src}(c, d + 1) + a\beta \cdot \text{src}(c + 1, d + 1), \quad (1)$$

where $\text{dst}(x, y)$ denotes the pixel value with coordinate (x, y) in the optical image, $\text{src}(c, d)$ denotes one of the four pixels closest

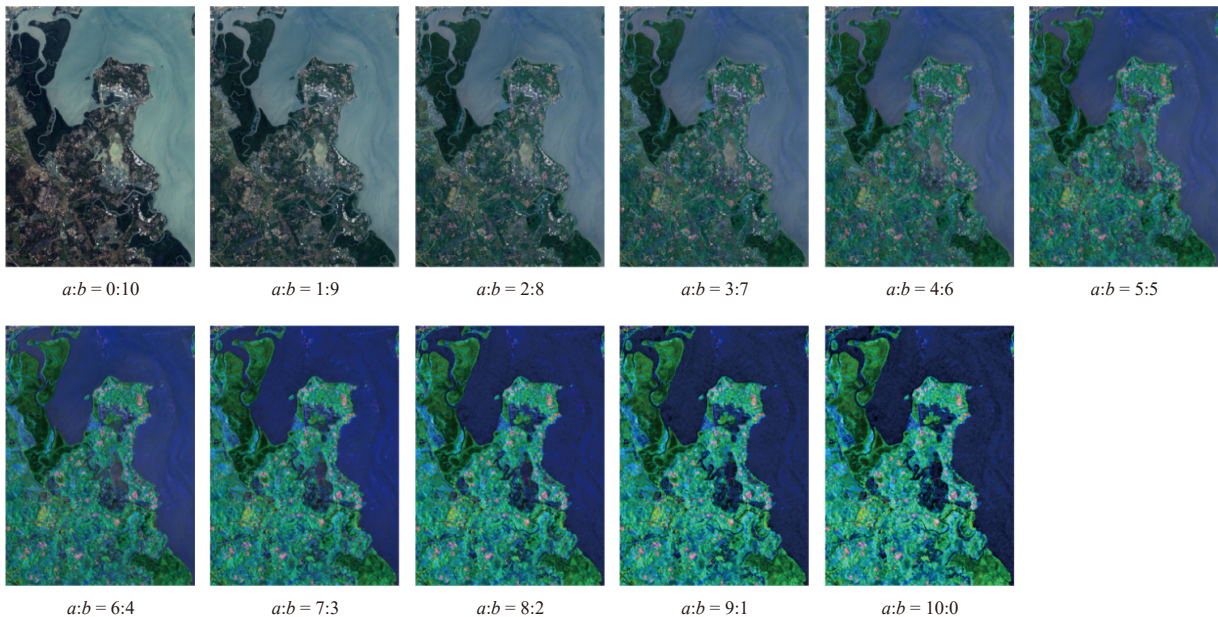


Fig. 2. Weighted fusion result of SAR image and optical image.

to (x, y) in the SAR image, α and β denote the fractional portion between coordinate (x, y) , and the nearest integer coordinate (c, d) , respectively.

Finally, two images with the same size and matching texture information are weighted and fused with the following formula:

$$\text{dst}(x, y) = \text{src1}(x, y) \cdot \alpha + \text{src2}(x, y) \cdot \beta + \gamma, \quad (2)$$

where $\text{dst}(x, y)$ denotes the pixel value with coordinate (x, y) in the output image, $\text{src1}(x, y)$ and $\text{src2}(x, y)$ denote the pixel values at the corresponding positions in the optical image and the SAR image, respectively, α and β denote the weights of the two images, i.e., a and b , respectively, and γ is used to adjust the brightness of the fused image.

4 Model and methods

4.1 AttU-Net neural network

The AttU-Net designed in this study was based on the framework of the U-Net network. The U-Net is a convolutional neural network with encoder and decoder parts for image segmentation. The encoder extracts advanced image features using a down-sampling operation, and the decoder restores the resolution using an upsampling operation. This type of structure can retain high-resolution information and effectively adapt to the complex structure and texture of mangrove vegetation.

However, the relative scarcity of training data and a more complex background environment should be addressed when researching mangrove vegetation identification. U-Net has potential issues in this context, mainly reflected in the following three aspects. First, U-Net is prone to overfitting as a deep neural network when training data are insufficient. Second, training can become unstable as the network deepens, leading to convergence, particularly when dealing with complex and diverse mangrove vegetation. Finally, because mangrove trees are similar to other trees, U-Net has a large difference in sensitivity to the input data for mangrove images under different environmental conditions, resulting in poor performance for mangrove scenes with large changes.

This paper improves the recognition performance of mangrove vegetation by adding a dropout layer, a batch normalization layer, and an attention mechanism to solve these problems. First, the dropout layer reduces the dependence between neurons by randomly dropping neurons during training, reducing overfitting and improving the model's generalization perform-

ance. Second, the BN layer. Mangrove vegetation varies under different conditions, such as light and humidity. The BN layer standardizes the middle layer's activation value, improves the network's robustness, and makes it suitable for the vegetation characteristics of different mangrove environments. Finally, this paper discusses attentional mechanisms. Mangrove vegetation exhibits complex structures and changes. An attention mechanism can make the network focus on areas more important for mangrove vegetation recognition, improving the network's perception of vegetation structure and texture and the recognition accuracy of mangrove vegetation.

Figure 3 shows the structure of the AttU-Net model (where S is the transition layer for attention-mechanism module processing, and D is the transition layer for the dropout operation).

In addition, because mangrove vegetation is a minority category in the overall image, the traditional cross-entropy loss function can overlearn the main category when dealing with unbalanced category distributions. Mangrove vegetation usually exists in coastal and marginal areas, often containing mangroves and other features. Their true labels are uncertain because edge pixels fall between two or more categories. Therefore, if the model focuses on identifying mangrove vegetation and prevents errors at the edge from being transmitted to the entire network through backpropagation, the convergence of the model will be affected. The AttU-Net network proposed adopts the improved ignoring edge cross-entropy function as the loss function and improves based on the classification cross-entropy function (CELoss). The parameter r is added to the denominator to adjust the size of the prediction region, and a weight is added to the molecule. Adjusting the weight allows the model to handle each category more balanced, improving the accuracy of mangrove vegetation identification. This paper denotes the improved neglected edge cross-entropy function as CLoss. The formula used is as follows:

$$\text{CLoss} = -\frac{1}{r \times N} \sum_{i=1}^{r \times N} \sum_{j=1}^G \omega_j y_{ij} \ln p_{ij}, \quad (3)$$

where N is the number of samples, G is the number of categories, y_{ij} indicates whether sample i belongs to category j , p_{ij} is the probability that sample i is predicted to be category j , and r is the ratio of pixels in the selected region to the number of pixels in the entire image to ignore the edge information, and ω_j is the weight of class j and adjusts the importance of class j .

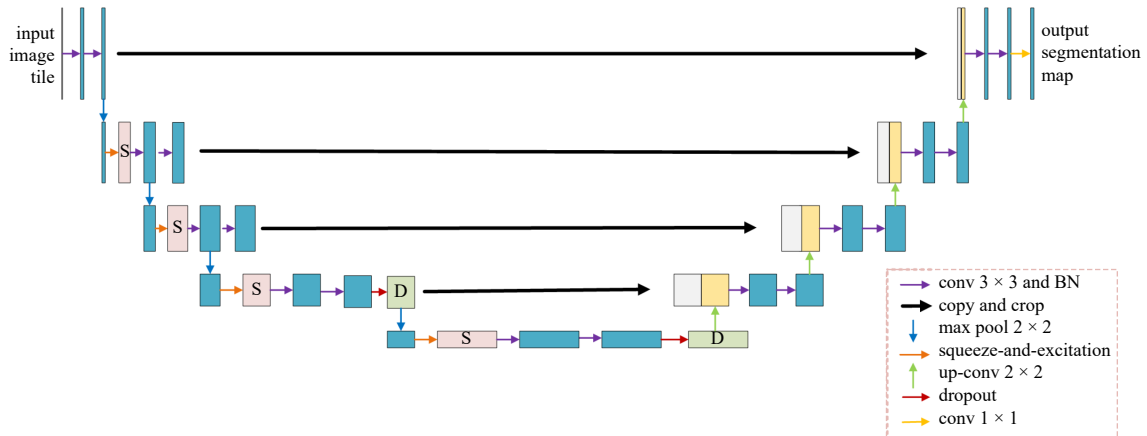


Fig. 3. Diagram of the AttU-Net model structure. Conv, convergence.

4.1.1 Dropout layer

The primary role of the dropout is to reduce overfitting and increase the model's generalization ability. The network learns more robust features by randomly "turning off" some neurons. Therefore, dropout is a simple and effective regularization method that improves the model's generalization performance. In mangrove vegetation recognition, owing to noise and complex environmental changes in the data, dropouts can effectively prevent the model from overfitting the training data and improve its adaptability to different mangrove scenes.

During training, dropout zeroes the neuron output with probability p by randomly "turning off" the neuron in each training iteration. The formula for dropout can be expressed as

$$\text{Dropout}(x) = \frac{\text{mask} \odot x}{1 - m}, \quad (4)$$

where x is the input feature, \odot represents the multiplication of element levels, and mask is a binary mask of the same size as the input, with each element zeroed with probability m .

4.1.2 Batch normalization layer

The BN layer normalizes each feature where the mean is close to zero, and its variance is close to 1. With learnable scaling and shifting parameters, the model can adapt to the characteristics of different distributions. The BN layer helps the network adapt better to different input distributions. This paper applied the BN layer before the activation function, preventing gradient explosion or disappearance, reducing the network's training time, and improving the model's generalization ability under limited samples.

The formula for batch normalization can be expressed as

$$\text{BN}(x) = \frac{A(x - \mu)}{\sigma} + B, \quad (5)$$

where x is the input feature, μ and σ are the mean and standard deviation of the input feature, respectively, and A and B are the learnable scaling and translation parameters.

4.1.3 Attention mechanism of squeeze-and-excitation network

Squeeze-and-excitation network (SE-Net) is a deep neural network based on an attention mechanism that improves a model's attention to the important features in the input. The core idea is to emphasize the importance of each channel in the network by learning adaptive weights to improve the model's representative ability.

It includes squeeze and excitation operations. The Squeeze phase, which enhances global pooling, compresses the information in each channel entering the feature map. Global average pooling obtained global information for each channel. The Excitation phase introduces two fully connected layers (FC layers) to learn the weights between channels. These two FC layers reduce

dimensionality (reducing the number of channels) and increase dimensionality (restoring the number of channels), and the weight is then generated using the sigmoid function. These weights are applied to the input feature map to obtain a weighted feature map. Figure 4 shows the structure of SE-Net, where F_{tr} is the traditional convolutional feature extraction structure, X and U are the inputs and outputs, respectively; F_{sq} is the Squeeze operation; F_{ex} is an Excitation operation; F_{scale} is the matrix multiplication between two channels; \tilde{X} is the result of F_{scale} processing; $X \in R^{H' \times W' \times C'}$, $U \in R^{H \times W \times C}$. The input feature graph $U \in R^{H \times W \times C}$, where H and W are the height and width of the feature graph, respectively, and C is the number of channels. The global descriptor $z \in R^C$ is obtained through global average pooling in the squeeze phase, representing each channel's global information. Two FC layers are introduced in the excitation phase, and z is mapped as the weight vector $s \in R^C$, where s_i represents the weight of the i channels. Finally, the weighted feature graph $Y \in R^{H \times W \times C}$ is obtained by multiplying the weight vector s with the original input feature graph U element by element, and the formula is as follows:

$$Y_i = s_i \times X_i. \quad (6)$$

Calculation formula of channel weight s_i as follows:

$$s_i = \sigma(W_2 \delta(W_1 z_i)), \quad (7)$$

where δ represents the ReLU activation function, σ is the sigmoid activation function, and W_1 and W_2 are the weight matrix of the FC layer.

Compared with other attention mechanisms, SE-Net focuses more on the adjustment of channel weights and emphasizes the importance of each channel, which is suitable for tasks that focus more on channel information. However, mangrove vegetation identification should emphasize specific channel information, such as the color and shape of the vegetation, and there are obvious differences between different channels. Therefore, SE-Net could better capture the information on these important channels by adjusting the channel weights.

4.2 Methods and procedures

The specific methods and steps of this study are as follows.

(1) Data processing: based on the fusion image, this paper combined the visual interpretation method with the vector data files measured in the field for manual annotation for true and accurate sample labels. The sample data set of 256 pixels \times 256 pixels is generated by sliding clipping. Then, the sample datasets obtained in this paper are divided into 3 000 training sets, 1 000 validation sets and 4 000 sample datasets. Among them, image mapping and labeling mapping are corresponded respectively.

(2) Construct the AttU-Net model: to improve the model's re-

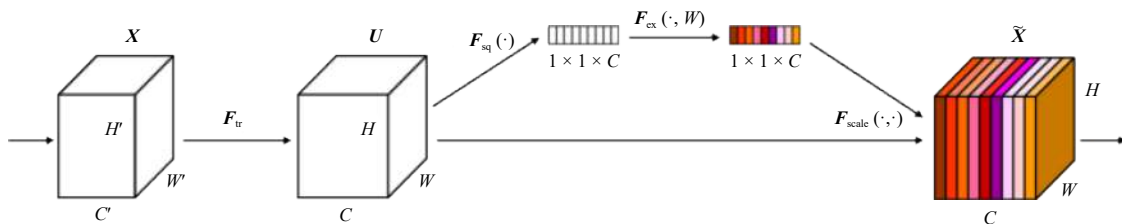


Fig. 4. Diagram of network structure of SE-Net.

cognition accuracy, make the model converge faster, and reduce overfitting of the background information. The U-Net network is enhanced to improve the recognition performance of mangrove vegetation by adding a dropout layer, BN layer, and attention mechanism.

(3) Training the AttU-Net model: in this paper, the AttU-Net model is trained based on a sample set of 11 fused images. Under the same parameter conditions, the fused image with the optimal accuracy evaluation is selected as the main study area. After that, the optimal mangrove vegetation recognition model in the changed study area is obtained by adjusting the parameters.

(4) Sliding splicing prediction: this paper introduces a sliding overlap splicing method to construct a prediction model. Its purpose is to be able to effectively eliminate the problem of splicing traces and enrich the edge information of the predicted image. The test set is inputted into the prediction model, and the prediction result map of network recognizing mangrove forest is obtained.

(5) Accuracy evaluation: the F1-score, overall accuracy (OA), and Kappa coefficient evaluated the mangrove vegetation classification results.

5 Experiments and results

5.1 Evaluation of experimental accuracy

The classification problem was a binary classification problem; this paper divided the image into mangrove and non-mangrove vegetation regions. In binary classification problems, a confusion matrix evaluates the performance of a classification model that provides a detailed classification of the model predictions about the actual class. The four main elements of the confusion matrix are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Table 3 presents the layout of the confusion matrix.

The TP correctly identified mangrove vegetation as vegetation among them. A TN occurs when the model correctly identifies a non-vegetated area as non-vegetated, a FP when the model incorrectly identifies non-vegetated areas as vegetation, and a FN when the model incorrectly identifies mangrove vegetation as non-vegetation.

Three evaluation factors were used to evaluate the mangrove vegetation classification results based on the confusion matrix. These were the F1-score, OA, and Kappa coefficient. The F1-score is an indicator that considers the precision and recall of the model, providing a single metric by balancing the model's performance in the positive and negative cases. The F1-score ranges from 0 to 1, with values closer to 1 indicating a better balance between accuracy and recall. The OA is a simple and intuitive evaluation indicator representing the proportion of the total sample number and the model correctly classifies across all categories. The Kappa coefficient measured the performance of the classification model. The range of the Kappa coefficient is between -1 and 1, and the closer it is to 1, the better the model's performance. Unlike the OA, Kappa coefficients are more robust for unbalanced categories and random guesses. The formulas for the F1-score, OA, and Kappa coefficient are as follows:

Table 3. Layout of confusion matrix

Prediction type	Real type	
	Terrace	Non-terraced field
Terrace	TP (true positive)	FP (false positive)
Non-terraced field	FN (false negative)	TN (true negative)

$$\text{F1-score} = 2 \times \frac{\text{TP}}{\text{TN} + \text{FP} + 2\text{TP}}, \quad (8)$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (9)$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e}. \quad (10)$$

The expressions for p_0 and p_e are as follows:

$$p_0 = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (11)$$

$$p_e = \frac{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) + (\text{FN} + \text{TN}) \times (\text{FP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2}. \quad (12)$$

5.2 Comparative experiments based on fusion images

In order to verify the effectiveness of the fused images proposed in this paper for mangrove vegetation recognition, the following comparison experiments are designed in this paper. The experiments are conducted based on 11 fused images weighted with different ratios, and the weighting ratio with optimal performance in the task of mangrove vegetation recognition is selected. The predicted area for the comparison experiment was a densely distributed area of mangrove vegetation with a size of 500 pixel \times 500 pixel. The accuracy evaluation indices were the F1-score, OA, and Kappa coefficient. Table 4 presents the accuracy evaluation results.

Figure 5 shows an experimental comparison figure of the comparison including the prediction image of the AttU-Net model trained based on fusion images of different proportions, the test image of fusion images of different proportions in the same region, and the real-label image of this region.

The experimental comparison chart shows the characteristics of the relatively concentrated areas of identification errors when the proportion of optical images is relatively high. This indicates that local details will likely affect the model when processing optical images. It is easy to confuse objects with similar colors or textures, resulting in the clustering of incorrect areas. In contrast, the experimental comparison graph shows a scattered dot distribution of the error areas in a high proportion of SAR images. Because SAR images can reveal the structure of ground objects, it is easier for the model to produce obvious boundaries between ground objects in the segmentation process; however, it

Table 4. Precision evaluation results of comparison experiments based on fusion images

Contrast region (a:b)	F1-score/%	OA/%	Kappa/%
0:10	96.696	94.347	77.192
1:9	98.282	97.016	86.968
2:8	98.567	97.506	88.959
3:7	96.688	94.350	77.542
4:6	96.067	93.349	74.674
5:5	97.643	95.936	82.919
6:4	96.039	93.257	73.462
7:3	98.067	96.598	83.897
8:2	96.969	94.721	76.504
9:1	96.438	93.835	73.547
10:0	95.635	92.481	68.569

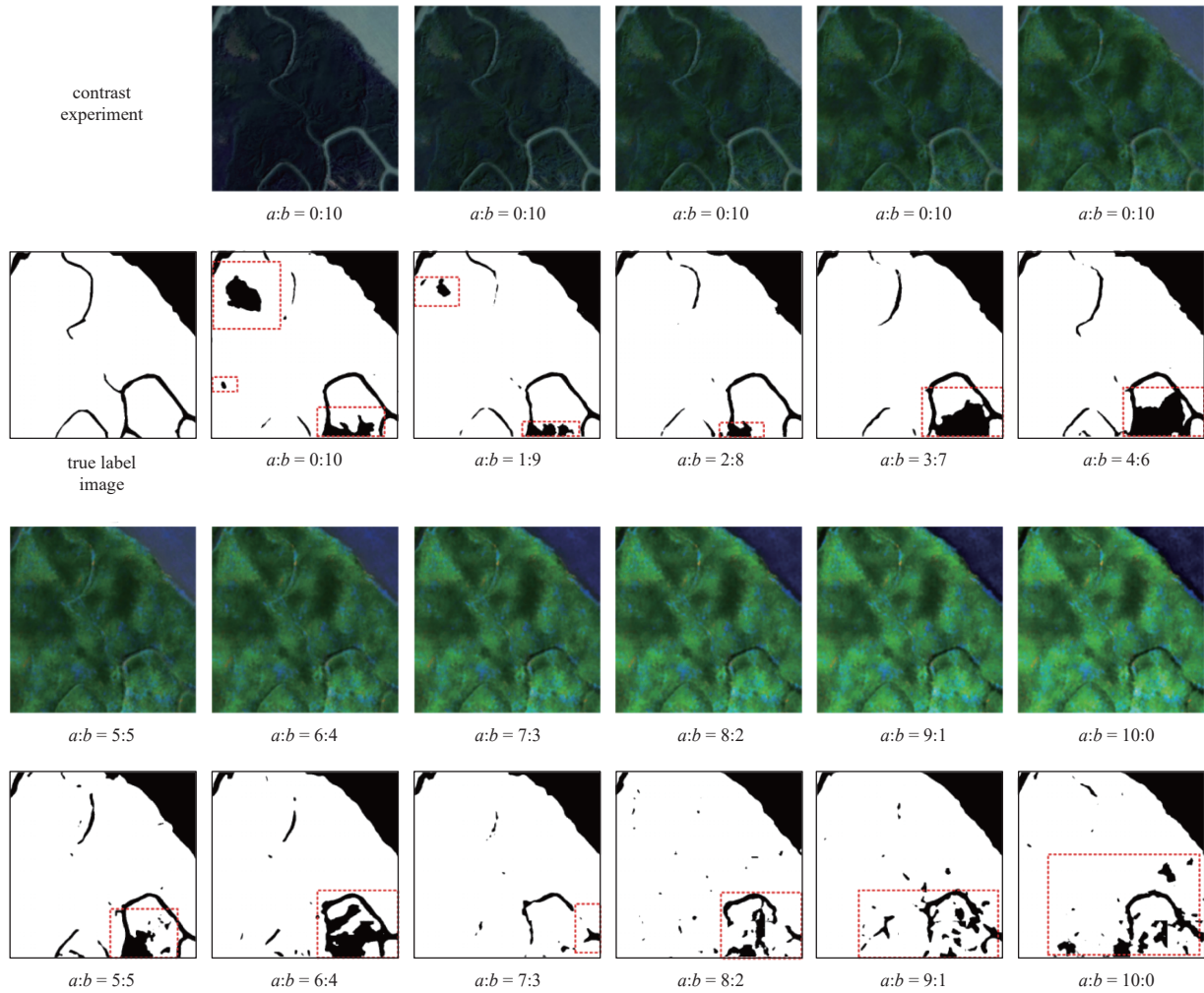


Fig. 5. Experimental comparison diagram of a comparison experiment. The white area in the image represents the area considered by the model as mangrove vegetation. The black area represents the area where the model considers non-mangrove vegetation. The red dashed box indicates the area with obvious recognition error.

is also easy to produce point-like misclassifications. Combined with the accuracy evaluation results of the comparative experiment, this paper selected the fusion proportion image with the best accuracy evaluation and best visual effect; that is, the main research object of the fusion image with $a:b = 2:8$. By introducing a smaller proportion of SAR image information, the model can better capture the boundary and structure of ground objects, avoiding the problem of overconcentration of identification errors in optical images.

5.3 Ablation experiment

Through the comparative experiments in Section 5.2, the fusion image with $a:b = 2:8$ is selected in this paper as the study image for the ablation experiments in this section. In order to better validate the effectiveness of the three modules introduced in this paper, ablation experiments are conducted on the SE-Net layer, Dropout layer and BN layer, respectively. As described in Section 5.2, the predicted area of the experiment was a densely distributed area of mangrove vegetation with 500 pixel \times 500 pixel. The accuracy evaluation indices were the F1-score, OA, and Kappa coefficient. This paper used the original U-Net network as the benchmark model for the experiment. Table 5 presents the accuracy evaluation results of the experiments. Figure 6 compares the ablation experiment's prediction results, including the test

image and the real label image of the mangrove vegetation.

First, according to the accuracy evaluation results of the ablation test area, compared with the baseline model, the model's OA, F1-scores, and Kappa coefficient significantly improved after adding the attention mechanism or dropout layer alone. This proves that an attention mechanism to make the model focus on the texture, structure, and details of specific areas improves the recognition of mangrove vegetation. In addition, by randomly

Table 5. Accuracy evaluation results of ablation test area

No.	Base	SE-Net	Drop.	BN	OA/%	F1-score/%	Kappa/%
1	√				95.038	97.107	79.694
2	√	√			96.948	98.239	86.797
3	√		√		96.697	98.092	85.814
4	√			√	93.804	96.352	75.891
5	√	√	√		95.755	97.529	82.494
6	√	√		√	94.962	97.073	79.015
7	√		√	√	95.203	97.206	80.302
8	√	√	√	√	97.507	98.568	88.959

Note: √ in the table proves that the module is added to the model; if √ is not marked, it proves that the module is not added to the model. Bold font denotes the highest value in this accuracy evaluation metric.

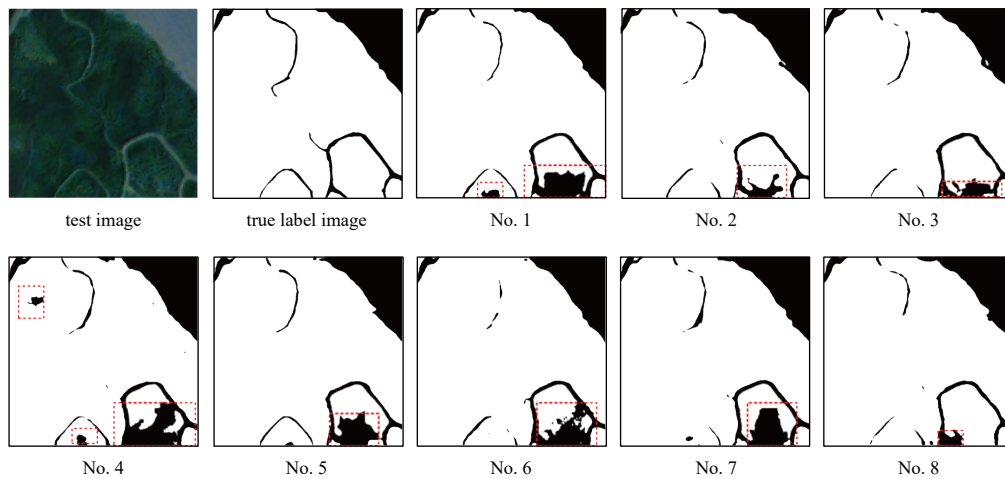


Fig. 6. Comparison of predicted results of ablation experiments. The white area in the image represents the area considered by the model as mangrove vegetation. The black area represents the area where the model considers non-mangrove vegetation. The red dashed box indicates the area with obvious recognition error.

discarding some neurons in the training process with a certain probability, overfitting the background information can be effectively reduced, and the model's generalization ability can be increased. However, after adding the BN layer alone, the OA, F1-scores, and Kappa coefficient decreased significantly compared with the benchmark model. Simultaneously, combined with the small figure No. 4 in Fig. 6, adding the BN layer alone introduces particular noise in the mangrove vegetation recognition task, resulting in the over-fitting of background information on the training set of the model. In the small figure No. 4 in Fig. 6, there is a segmentation error in the upper left corner that does not appear in other predictions. This is because the U-Net model is not complicated, and the expression ability of the model is insufficient after the addition of the BN layer, which decreases the model's performance.

In the models with both modules added simultaneously, the accuracy of the models decreases (No. 5, No. 6, and No. 7 in Fig. 6) compared with the models that add the attention mechanism or dropout layer alone (No. 2 and No. 3 in Fig. 6). Compared with the model with the BN layer alone (No. 4), the accuracy of the model is improved (No. 6 and No. 7 in Fig. 6). Combined with the images, the simultaneous addition of the attention mechanism and dropout layer introduces complexity to the model, making the feature distribution more dynamic and elusive. As a result, overfitting of the background information occurred in the model. As shown in Fig. 5, the area of the identification error in the lower-right corner was significantly larger. Based on the model that had already added the BN layer, the addition of the attention mechanism and dropout layer improved (No. 6 and No. 7 in Fig. 6). Compared to any other model, the F1-score, OA, and Kappa coefficient of the model with the last three modules added simultaneously were the best. Combined with the small figure No. 8 in Fig. 6, the identification error area was the smallest, and there were no other situations in which mangrove areas were identified as non-mangrove areas. Thus, the attention mechanism and dropout layer introduce complexity to the model, making the feature distribution more dynamic and elusive. However, the BN layer normalizes the distribution of features in a complex model, making it easier for the model to converge and improving model recognition accuracy. In summary, adding SE-Net, dropout, and BN modules can simultaneously improve the water-body recognition ability of the model.

5.4 Comparative experiment

To verify the model's performance more comprehensively and compare it with its benchmark model U-Net and other mainstream Seg-Net, Dense-Net, and Res-Net deep learning networks, four areas outside the sample area were selected for prediction in this paper. The sizes of the areas are both 500 pixels \times 500 pixels. The parameter settings of the model are shown in Table 6.

Figure 7 shows the accuracy and loss of the training and test sets based on the fusion images of the model used in this study. The line graph on the left represents the accuracy of the training and verification sets, with the horizontal coordinate representing the number of iterations and the vertical coordinate representing the accuracy. The line chart on the right represents the loss values of the training and verification sets, where the horizontal coordinate is the number of iterations and the vertical coordinate is the loss value.

The prediction results of AttU-Net, the training model network of this paper, are compared with its benchmark model U-Net as well as three other mainstream deep learning networks, Seg-Net, Dense-Net, and Res-Net, for the four selected test regions. The evaluation metrics are F1-score, OA, and Kappa coefficient. The comparison of the prediction results is shown in Fig. 8. The accuracy evaluation results for test regions 1–4 are shown in Table 7.

Figure 8 shows that the U-Net network presents a better visual effect than the other three mainstream deep-learning networks: Seg-Net, Dense-Net, and Res-Net. The AttU-Net network proposed inherits the characteristics of the U-Net network structure to retain high-resolution information, has the best visual results, more accurate identification of mangrove vegetation, and better ability to identify mangrove and non-mangrove areas. It can adapt more effectively to the complex structure and texture of mangrove vegetation.

According to the accuracy evaluation results in Table 7, the AttU-Net model proposed in this paper outperforms other mod-

Table 6. Training parameter settings

Parameter	Specific setting
Batch size	16
Learning rate	1×10^{-4}
Epoch	65
Optimizer	Adam

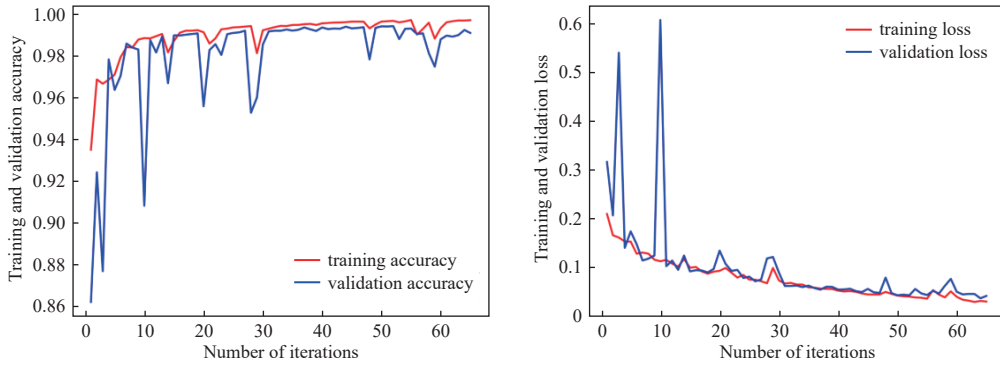


Fig. 7. Line charts of accuracy and loss values of the training set and test set based on fusion images.

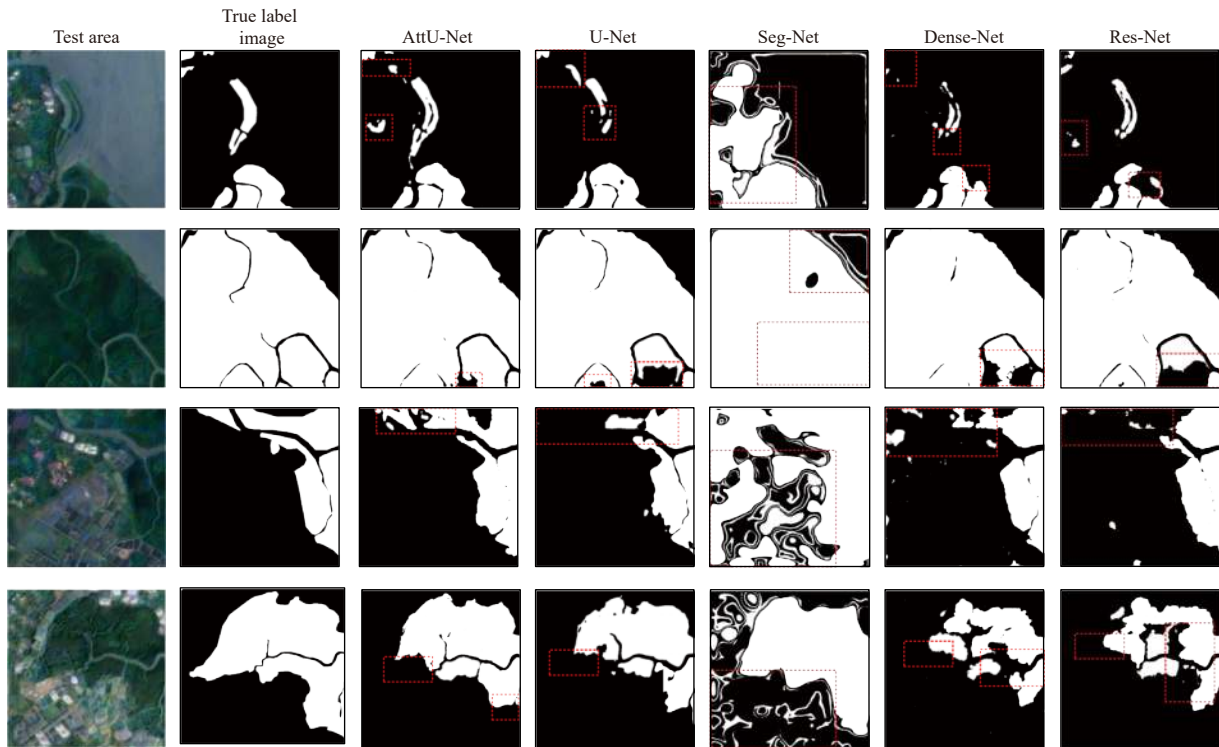


Fig. 8. Comparison of prediction results of mangrove vegetation in the test area. The white area in the image represents the area considered by the model as mangrove vegetation. The black area represents the area where the model considers non-mangrove vegetation. The red dashed box indicates the area with obvious recognition error.

els. Bold font denotes the highest value in this accuracy evaluation metric.

In test area 1, characterized by fewer mangrove areas, the AttU-Net model demonstrates a substantial improvement in the F1-score and Kappa coefficient compared to the benchmark network and three other networks. Specifically, the F1-score and Kappa coefficient of AttU-Net increased by 6.425% and 7.08% respectively compared to the benchmark U-Net model. Additionally, its F1-score is 12.906% higher than that of Res-Net, the best-performing model among the other three in terms of F1-score, and its Kappa coefficient is 13.903% higher than that of Dense-Net, which had the highest Kappa coefficient among the other models. Hence, in areas with fewer mangroves, AttU-Net surpasses the performance of other networks.

In test area 2, although the overall accuracy and F1-scores of AttU-Net were not significantly different from those of the other models, its Kappa coefficient was superior, reaching 88.959%.

This suggests that while other networks achieve high accuracy and comprehensive performance in mangrove vegetation recognition, they fall short in terms of consistency and randomness of classification. Figure 8 illustrates that Seg-Net’s prediction in test area 2 exhibits noticeable flaws, particularly in the background details. The river identification is either poor or entirely absent, resulting in a Kappa coefficient of only 58.229%, despite an OA of 92.363% and an F1-score of 95.753%. This indicates difficulty in maintaining classification consistency across different categories in imbalanced classes. The AttU-Net model, however, displayed the best Kappa coefficient, with an improvement of 9.265% over U-Net and 8.034% over Dense-Net, the top performer among the comparison models.

In test area 3, affected by farmland and house interference, all models experienced a performance decline, and the overall accuracy did not reach the average levels of the first two areas. This suggests that human interference likely alters the characteristics

Table 7. Accuracy evaluation results of test area

Test area	Model	Accuracy evaluation		
		OA/%	F1-Score/%	Kappa/%
Test area 1	AttU-Net (ours)	97.082	88.008	86.348
	U-Net	95.870	81.583	79.268
	Seg-Net	71.099	39.136	25.900
	Dense-Net	95.056	75.064	72.445
	Res-Net	94.974	75.102	72.410
Test area 2	AttU-Net (ours)	97.506	98.567	88.959
	U-Net	95.038	97.107	79.694
	Seg-Net	92.363	95.753	58.229
	Dense-Net	94.571	96.835	80.925
	Res-Net	94.083	96.524	76.728
Test area 3	AttU-Net (ours)	93.952	87.878	83.851
	U-Net	93.553	86.171	82.009
	Seg-Net	51.064	50.002	19.944
	Dense-Net	91.625	82.041	76.633
	Res-Net	92.383	83.158	78.328
Test area 4	AttU-Net (ours)	89.083	85.572	77.021
	U-Net	85.762	80.093	69.644
	Seg-Net	83.485	82.329	67.046
	Dense-Net	78.289	65.889	52.542
	Res-Net	80.267	69.966	57.147

Note: Bold font denotes the highest value in this accuracy evaluation metric.

of mangrove forests, such as texture, shape, and color, making accurate identification more difficult. In areas adjacent to houses, the texture and shape of the mangrove regions changed more significantly, and the color became lighter. The AttU-Net model showed improvements over U-Net, the best-performing benchmark, with increases in OA, F1-score, and Kappa coefficient by 0.399%, 1.707%, and 1.842%, respectively.

In test area 4, which experiences more significant interference from farmland and houses, none of the four models achieved an overall accuracy exceeding 90%. The performance of all models, except Seg-Net, deteriorated significantly. However, the proposed AttU-Net model maintained relatively high performance, with the Kappa coefficient improving by 7.377% compared to U-Net, the model with the highest Kappa coefficient among the other three models. Additionally, the F1-score improved by 3.243% compared to Seg-Net, and the overall accuracy increased by 3.321% compared to U-Net. In this test area, Seg-Net's performance showed a notable improvement compared to test area 3. Analyzing results across test areas 1–4 reveals that the Seg-Net model is particularly sensitive to green color and zigzag texture patterns; it performs better with larger proportions of mangrove areas and worse with smaller proportions.

In summary, by comparing the mangrove vegetation prediction results across the test areas and the accuracy evaluation results from test areas 1–4, AttU-Net demonstrated higher overall performance, better detail capture ability, and greater robustness against category imbalance in mangrove vegetation identification tasks. Therefore, AttU-Net is an effective model for the high-precision identification of mangrove vegetation in fusion images and can significantly contribute to the monitoring and protection of mangrove ecosystems.

6 Conclusions

This paper proposes a pixel-level weighted fusion method for SAR and optical images to extract mangrove vegetation information more accurately. At the method level, an AttU-Net model

was established to identify mangrove vegetation accurately. To verify the effectiveness of the fusion image, this study employed the AttU-Net model and trained it using various weighted ratios of the fusion image. Ultimately, through comparative experimentation, a weighted ratio of 2:8 for the fusion image was selected as the most effective. To verify the validity of the AttU-Net model, it was compared with the prediction results of the benchmark model U-Net and three other mainstream deep-learning networks, Seg-Net, Dense-Net, and Res-Net, for the two selected test areas. The results showed that the model had higher overall performance, better detail capture ability, and better robustness against category imbalance in identifying mangrove vegetation, with average OA, F1-score, and Kappa coefficients of 94.406%, 90.006%, and 84.045% in the four test areas, respectively. Prove that this method can play a positive role in monitoring and protecting mangrove vegetation.

References

- Braun A C. 2021. More accurate less meaningful? A critical physical geographer's reflection on interpreting remote sensing land-use analyses. *Progress in Physical Geography: Earth and Environment*, 45(5): 706–735, doi: [10.1177/0309133321991814](https://doi.org/10.1177/0309133321991814)
- Cao Jingjing, Leng Wanchun, Liu Kai, et al. 2018. Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. *Remote Sensing*, 10(1): 89, doi: [10.3390/rs10010089](https://doi.org/10.3390/rs10010089)
- Chen Zhaojun, Zhang Meng, Zhang Huaqing, et al. 2023. Mapping mangrove using a red-edge mangrove index (REMI) based on Sentinel-2 multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4409511
- Darko P O, Kalacska M, Arroyo-Mora J P, et al. 2021. Spectral complexity of hyperspectral images: A new approach for mangrove classification. *Remote Sensing*, 13(13): 2604, doi: [10.3390/rs13132604](https://doi.org/10.3390/rs13132604)
- de Souza Moreno G M, de Carvalho Júnior O A, de Carvalho O L F, et al. 2023. Deep semantic segmentation of mangroves in Brazil combining spatial, temporal, and polarization data from Sentinel-1 time series. *Ocean & Coastal Management*, 231: 106381
- Fu Bolin, Liang Yiyin, Lao Zhinan, et al. 2023. Quantifying scattering characteristics of mangrove species from Optuna-based optimal machine learning classification using multi-scale feature selection and SAR image time series. *International Journal of Applied Earth Observation and Geoinformation*, 122: 103446, doi: [10.1016/j.jag.2023.103446](https://doi.org/10.1016/j.jag.2023.103446)
- Fu Chang, Song Xiqiang, Xie Yu, et al. 2022. Research on the spatiotemporal evolution of mangrove forests in the Hainan Island from 1991 to 2021 based on SVM and Res-UNet Algorithms. *Remote Sensing*, 14(21): 5554, doi: [10.3390/rs14215554](https://doi.org/10.3390/rs14215554)
- Giri C. 2016. Observation and monitoring of mangrove forests using remote sensing: opportunities and challenges. *Remote Sensing*, 8(9): 783, doi: [10.3390/rs8090783](https://doi.org/10.3390/rs8090783)
- Gonzalez-Perez A, Abd-Elrahman A, Wilkinson B, et al. 2022. Deep and machine learning image classification of coastal wetlands using unpiloted aircraft system multispectral images and Lidar datasets. *Remote Sensing*, 14(16): 3937, doi: [10.3390/rs14163937](https://doi.org/10.3390/rs14163937)
- Huang Sha, Tang Lina, Hupy J P, et al. 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1): 1–6, doi: [10.1007/s11676-020-01155-1](https://doi.org/10.1007/s11676-020-01155-1)
- Jia Mingming, Wang Zongming, Wang Chao, et al. 2019. A new vegetation index to detect periodically submerged mangrove forest using single-tide Sentinel-2 imagery. *Remote Sensing*, 11(17): 2043, doi: [10.3390/rs11172043](https://doi.org/10.3390/rs11172043)
- Kamal M, Phinn S, Johansen K. 2014. Characterizing the spatial structure of mangrove features for optimizing image-based mangrove mapping. *Remote Sensing*, 6(2): 984–1006, doi: [10.3390/rs6020984](https://doi.org/10.3390/rs6020984)
- Kulkarni S C, Rege P P. 2020. Pixel level fusion techniques for SAR

- and optical images: a review. *Information Fusion*, 59: 13–29, doi: [10.1016/j.inffus.2020.01.003](https://doi.org/10.1016/j.inffus.2020.01.003)
- Li Jinjin, Zhang Jiacheng, Yang Chao, et al. 2023. Comparative analysis of pixel-level fusion algorithms and a new high-resolution dataset for SAR and optical image fusion. *Remote Sensing*, 15(23): 5514, doi: [10.3390/rs15235514](https://doi.org/10.3390/rs15235514)
- Lu Ying, Wang Le. 2021. How to automate timely large-scale mangrove mapping with remote sensing. *Remote Sensing of Environment*, 264: 112584, doi: [10.1016/j.rse.2021.112584](https://doi.org/10.1016/j.rse.2021.112584)
- Luo Yanmin, Ouyang Yi, Zhang Rencheng, et al. 2017. Multi-feature joint sparse model for the classification of mangrove remote sensing images. *ISPRS International Journal of Geo-Information*, 6(6): 177, doi: [10.3390/ijgi6060177](https://doi.org/10.3390/ijgi6060177)
- Mahmoud M I. 2012. Information extraction from paper maps using object oriented analysis (OOA) [dissertation]. Enschede: University of Twente
- Maurya K, Mahajan S, Chaube N. 2021. Remote sensing techniques: mapping and monitoring of mangrove ecosystem—A review. *Complex & Intelligent Systems*, 7(6): 2797–2818
- Purnamasayangasukasih P R, Norizah K, Ismail A A M, et al. 2016. A review of uses of satellite imagery in monitoring mangrove forests. *IOP Conference Series: Earth and Environmental Science*, 37: 012034, doi: [10.1088/1755-1315/37/1/012034](https://doi.org/10.1088/1755-1315/37/1/012034)
- Raghavendra N S, Deka P C. 2014. Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing*, 19: 372–386, doi: [10.1016/j.asoc.2014.02.002](https://doi.org/10.1016/j.asoc.2014.02.002)
- Sandra M C, Rajitha K. 2023. Random forest and support vector machine classifiers for coastal wetland characterization using the combination of features derived from optical data and synthetic aperture radar dataset. *Journal of Water & Climate Change*, 15(1): 29–49
- Shen Zhen, Miao Jing, Wang Junjie, et al. 2023. Evaluating feature selection methods and machine learning algorithms for mapping mangrove forests using optical and synthetic aperture radar data. *Remote Sensing*, 15(23): 5621, doi: [10.3390/rs15235621](https://doi.org/10.3390/rs15235621)
- Su Jiming, Zhang Fupeng, Yu Chuanxiu, et al. 2023. Machine learning: next promising trend for microplastics study. *Journal of Environmental Management*, 344: 118756, doi: [10.1016/j.jenvman.2023.118756](https://doi.org/10.1016/j.jenvman.2023.118756)
- Tian Lei, Wu Xiaocan, Tao Yu, et al. 2023. Review of remote sensing-based methods for forest aboveground biomass estimation: progress, challenges, and prospects. *Forests*, 14(6): 1086, doi: [10.3390/f14061086](https://doi.org/10.3390/f14061086)
- Toosi N B, Soffianian A R, Fakheran S, et al. 2019. Comparing different classification algorithms for monitoring mangrove cover changes in southern Iran. *Global Ecology and Conservation*, 19: e00662., doi: [10.1016/j.gecco.2019.e00662](https://doi.org/10.1016/j.gecco.2019.e00662)
- Tran T V, Reef R, Zhu Xuan. 2022. A review of spectral indices for mangrove remote sensing. *Remote Sensing*, 14(19): 4868, doi: [10.3390/rs14194868](https://doi.org/10.3390/rs14194868)
- Twilley R R. 2019. Mangrove wetlands. In: Messina M G, Conner W H, eds. *Southern Forested Wetlands*. London: Routledge, 445–473
- Wang Pin, Fan En, Wang Peng. 2021a. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141: 61–67, doi: [10.1016/j.patrec.2020.07.042](https://doi.org/10.1016/j.patrec.2020.07.042)
- Wang Youshao, Gu Jidong. 2021b. Ecological responses, adaptation and mechanisms of mangrove wetland ecosystem to global climate change and anthropogenic activities. *International Biodegradation & Biodegradation*, 162: 105248
- Wei Yidi, Cheng Yongcun, Yin Xiaobin, et al. 2023. Deep learning-based classification of high-resolution satellite images for mangrove mapping. *Applied Sciences*, 13(14): 8526, doi: [10.3390/app13148526](https://doi.org/10.3390/app13148526)
- Xie Yiheng, Chen Renxi, Yu Mingge, et al. 2023. Improvement and application of UNet network for avoiding the effect of urban dense high-rise buildings and other feature shadows on water body extraction. *International Journal of Remote Sensing*, 44(12): 3861–3891, doi: [10.1080/01431161.2023.2229498](https://doi.org/10.1080/01431161.2023.2229498)
- Xu Chen, Wang Juanle, Sang Yu, et al. 2023a. An effective deep learning model for monitoring mangroves: a case study of the Indus delta. *Remote Sensing*, 15(9): 2220, doi: [10.3390/rs15092220](https://doi.org/10.3390/rs15092220)
- Xu Mengjie, Sun Chuanwang, Zhan Yanhong, et al. 2023b. Impact and prediction of pollutant on mangrove and carbon stocks: a machine learning study based on urban remote sensing data. *Geoscience Frontiers*, 15(3): 101665
- Yang Gang, Huang Ke, Sun Weiwei, et al. 2022. Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. *ISPRS Journal of Photogrammetry and Remote Sensing*, 189: 236–254, doi: [10.1016/j.isprsjprs.2022.05.003](https://doi.org/10.1016/j.isprsjprs.2022.05.003)
- Yu Mingge, Rui Xiaoping, Zou Yarong, et al. 2023. Research on automatic recognition of mangrove forests based on CU net model. *Journal of Oceanography (in Chinese)*, 45(3): 125–135
- Zhang Junyao, Yang Xiaomei, Wang Zhihua, et al. 2021. Remote sensing based spatial-temporal monitoring of the changes in coastline mangrove forests in China over the last 40 years. *Remote Sensing*, 13(10): 1986, doi: [10.3390/rs13101986](https://doi.org/10.3390/rs13101986)