

Article

Not peer-reviewed version

An Approach for Detecting Mangrove Areas and Mapping Species Using Multispectral Drone Imagery and Deep Learning

[Xingyu Chen](#) , [Xiuyu Zhang](#) ^{*} , [Changwei Zhuang](#) , [Xuejiao Dai](#) , Lingling Kong , Zixia Xie , Xibang Hu

Posted Date: 17 February 2025

doi: 10.20944/preprints202502.1120.v1

Keywords: Mangrove ecosystems; high-resolution multispectral data; mangrove species identification; convolutional neural network; spectral and spatial attention mechanisms



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

An Approach for Detecting Mangrove Areas and Mapping Species Using Multispectral Drone Imagery and Deep Learning

Xingyu Chen ^{1,2}, Xiuyu Zhang ^{1,2,*}, Changwei Zhuang ^{1,2}, Xuejiao Dai ^{1,2}, Lingling Kong ^{1,2},
Xixia Xie ^{1,2} and Xibang Hu ^{1,2}

¹ Institute of Ecological Civilization and Green Development, Guangdong Provincial Academy of Environmental Science, Guangzhou 510045, China; chenxy2019@mail.sustech.edu.cn (X.C.); ecozcw@163.com (C.Z.); 869914547@qq.com (X.D.); kll_linxi1009@163.com (L.K.); xiezx.18s@igsrr.ac.cn (Z.X.); hooyan@126.com (X.H.)

² Ecological Environment Remote Sensing Research Center, Guangdong Provincial Academy of Environmental Science, Guangzhou 510045, China

* Correspondence: zhangxiuyu@scies.org

Abstract: Mangrove ecosystems are important in tropical and subtropical coastal zones, contributing to marine biodiversity and maintaining marine ecological balance. It is crucial to develop more efficient, intelligent, and accurate monitoring methods for mangroves to understand better and protect mangrove ecosystems. This study promoted a novel model, MangroveNet, for integrating multi-scale spectral and spatial information and detecting mangrove area. In addition, we also presented an improved model—AttCloudNet+ to identify distribution of mangroves species based on high-resolution multispectral drone images. These models incorporate spectral and spatial attention mechanisms and have been shown to effectively address the limitations of traditional methods, which have been prone to inaccuracy and low efficiency in mangrove species identification. In this study, we compare the results from MangroveNet with the SegNet, UNet, and DeepUNet models. The findings demonstrate that the MangroveNet exhibits superior generalization learning capabilities and more accurate extraction outcomes than other deep learning models. The accuracy, F1_Score, mIoU, and precision of MangroveNet were 99.13%, 98.84%, 98.11%, and 99.14%, respectively. In terms of identifying mangrove species, the prediction results from AttCloudNet+ were compared with those obtained from traditional supervised and unsupervised classifications and various machine learning and deep learning methods. These included K-means clustering, ISODATA cluster analysis, random forest (RF), support vector machines (SVM), and others. The comparison demonstrated that the mangrove species identification results obtained using AttCloudNet+ exhibited the most optimal performance in terms of the Kappa coefficient and the overall accuracy (OA) index, reaching 0.81 and 0.87, respectively. The two comparison results confirm the effectiveness of the two models developed in this study for identifying mangroves and their species. Overall, we provided an efficient solution based on deep learning with a dual attention mechanism in the acceptable real-time monitoring of mangroves and their species using high-resolution multispectral drone imagery.

Keywords: Mangrove ecosystems; high-resolution multispectral data; mangrove species identification; convolutional neural network; spectral and spatial attention mechanisms

1. Introduction

Mangroves are one of the most sensitive parts of the marine ecosystem to ecological environment change[1]. The variation of mangroves and their species can assist in understanding the health status of mangrove ecosystems, biodiversity, ecosystem function, and their response to climate change[2,3].

Given the extensive distribution of mangroves across the global coastal zone[4], monitoring the range and internal species distribution of mangroves with high spatial and temporal accuracy is of great significance for analyzing the changes in regional and even global marine ecosystems, including those resulting from climate, ecological environment and other factors[5,6].

The initial ecological monitoring of mangrove ecosystems predominantly employed artificial field sampling surveys to ascertain approximate mangrove ranges and internal species numbers[7]. However, this approach proved inadequate for determining the spatial distribution of mangrove species [8]. Furthermore, this approach necessitates significant financial investment and substantial human and temporal resources, while monitoring efficiency remains challenging. The advent of satellite remote sensing data saw researchers begin to analyze and obtain the ecological status of mangroves from remote sensing imagery[4,5,7,9,10]. However, high-resolution optical remote sensing images are not as readily accessible as public medium-resolution remote sensing images[11-13], (e.g., Landsat series and Synthetic Aperture Radar from Sentinel Satellite)[4].

Furthermore, optical remote sensing images are susceptible to interference from cloud and precipitation, which further reduces the number of available optical images[14]. Drone imagery data can effectively address this limitation, as drones can fly below the cloud cover, obtaining ultra-high-resolution images without being affected by cloudy weather conditions[15,16]. Additionally, the multispectral frequency band information provided by satellite images is frequently employed to differentiate the characteristics of disparate ground objects[17,18]. Nevertheless, despite the availability of various spectral bands in satellite image data, the low resolution of such images still presents a challenge in distinguishing between mangrove species[19]. To address this issue, we employ multispectral sensors mounted on drones to enhance the conventional color images drones capture into multi-spectral, high-resolution, spatial images encompassing visible light bands, red-edge band and near-infrared band[19,20].

Regarding the mangrove extraction algorithm, research has been conducted on utilizing a mangrove index based on remote sensing images to identify and obtain an accurate delineation of the mangrove range[21]. The method of mangrove information extraction based on spectral index has been demonstrated to offer high efficiency and broad applicability. In addition to investigating the potential applications of conventional spectral indices in mangrove information extraction, researchers have also attempted to construct mangrove indices based on the distinctive spectral properties of mangrove vegetation[22,23]. Shi et al. (2016)[24] proposed the Normalized Difference Mangrove Index (NDMI) based on Landsat-8 OLI images as a means of enhancing the separability of mangrove and terrestrial vegetation. Diniz et al. (2019)[25] proposed the use of the Modular Mangrove Recognition Index (MMRI) based on Landsat TOA images for the analysis of dynamic changes in Brazilian mangroves from 1985 to 2018. Huete et al. (2002)[26] combined Landsat-5 TM, Landsat-7 ETM+ and Landsat-8 OLI images to construct and combined the Enhanced Vegetation Index (EVI), Modified Soil-Adjusted Vegetation Index (MSA-VI)[27], Normalized Difference Vegetation Index (NDVI)[28] and NDMI to analyze the health of mangrove forests in a coastal lagoon.

The methods above are relatively straightforward to implement. However, the robustness of these methods is still being determined, given the necessity for applying diverse thresholds[29,30] in the context of images captured under disparate conditions, including those about time and sun angle. To address this challenge, data-driven machine learning offers an alternative approach[31]. For instance, researchers have employed machine learning algorithms, such as Random Forest (RF)[32], Support Vector Machine (SVM)[33], and Artificial Neural Networks (ANNs)[34], to extract mangroves from remote sensing images. However, the efficacy of machine learning based algorithms hinges on expertise, and the performance of the corresponding methodology is contingent on the quality of the handcrafted feature selection[31].

Deep learning algorithms have demonstrated remarkable efficacy in many remote sensing interpretation and image segmentation scenarios[35]. The extensive utilization of deep learning algorithms in remote sensing interpretation has led to their emergence as a superior alternative to traditional remote sensing algorithms and conventional machine learning algorithms. It can autonomously learn high-level semantic information in images, which enhances segmentation accuracy. This capability has led to a growing use of deep learning algorithms in remote sensing

image processing, particularly for identifying and classifying mangrove ecosystems[36]. In a study conducted by Dang et al. (2020)[37], the performance of the U-Net model and SVM in classifying mangrove forests was compared. The findings revealed that the U-Net model demonstrated superior performance, achieving an accuracy rate of 90% and a Kappa coefficient value of 0.84. In comparison, the SVM attained an accuracy of 86% and a Kappa coefficient of 0.78[37]. Faza et al. (2018)[38] built a deep neural network architecture and analyzed it for mangrove sprout plant classification. The result from the architectural methodology of this research reached the lowest training error of 0.1345 and the highest testing accuracy value of 98%[38].

The following section outlines the structure of the paper. Section 2 presented an overview of the study region and the datasets employed. Section 3 provided a detailed account of the architectural design of the MangroveNet and AttCloudNet+, the data processing methodology, the extraction of mangrove areas, and the identification of mangrove species using our deep learning models. Section 4 presented the precision indices performance and model validation of the MangroveNet and other CNN segmentation models. Furthermore, the AttCloudNet+ was used to categorize mangrove species and evaluate their performance compared to other traditional methods and machine learning algorithms (e.g., ISODATA, Kmeans[39], RF, SVM, etc.). In the discussion section, factors affecting the results of each model and the effect of shadow cover on the mangrove extraction misclassification in drone images were demonstrated. The conclusion is described in Section 6. This study is highlighted by developing modified deep-learning architectures to study mangroves and their species identification via drone multispectral high-resolution imagery.

2. Study Area and Data

2.1. Study Area

The selected research area is situated within the Guangdong Zhanjiang Mangrove National Nature Reserve's largest mangrove national nature reserve in China[40]. The reserve is geo-located on the Leizhou Peninsula of Guangdong Province, occupying a geographically distinctive position at the southernmost extremity of the Chinese mainland[41]. The coordinates of this region are between 109°40' and 110°35' east longitude and 20°14' and 21°35' north latitude[41]. The remainder of the area is characterized by terraces and plains with gentle slopes[42]. The highest altitude of the peninsula is 382 meters, and the region is devoid of elevated mountainous terrain. The mean annual temperature of the reserve is approximately 23 °C, with a maximum temperature of 38.8 °C and a minimum temperature of -1.4 °C[43]. The average annual precipitation in the area is 1534.6 mm, indicating a clear distinction between wet and dry seasons.

This study acquired in-situ drone image data near Jinniu Island in the Huguang Jinxing partition, situated on the eastern Zhanjiang Mangrove National Nature Reserve, as shown in **Error! Reference source not found.** The ecosystem in this region is relatively complete and possesses high protection and scientific research value[44]. A comprehensive scientific investigation report and research in this reserve revealed the existence of nine basic mangrove community types within the reserve[45]. These include the *Avicennia marina* group association, the *Rhizophora stylosa* group association, the *Aegiceras corniculatum* group association, the *Sonneratia apetala* group association, the *Kandelia obovata* group association, the *Bruguiera gymnorrhiza* group association, the *Acrostichum aureum* group association, the *Acanthus ilicifolius* group association, as well as the *Excoecaria agallocha* group association[46]. The most prevalent mangrove species near Jinniu Island are *Rhizophora stylosa*, *Aegiceras corniculatum*, *Sonneratia apetala*, and *Avicennia marina*.

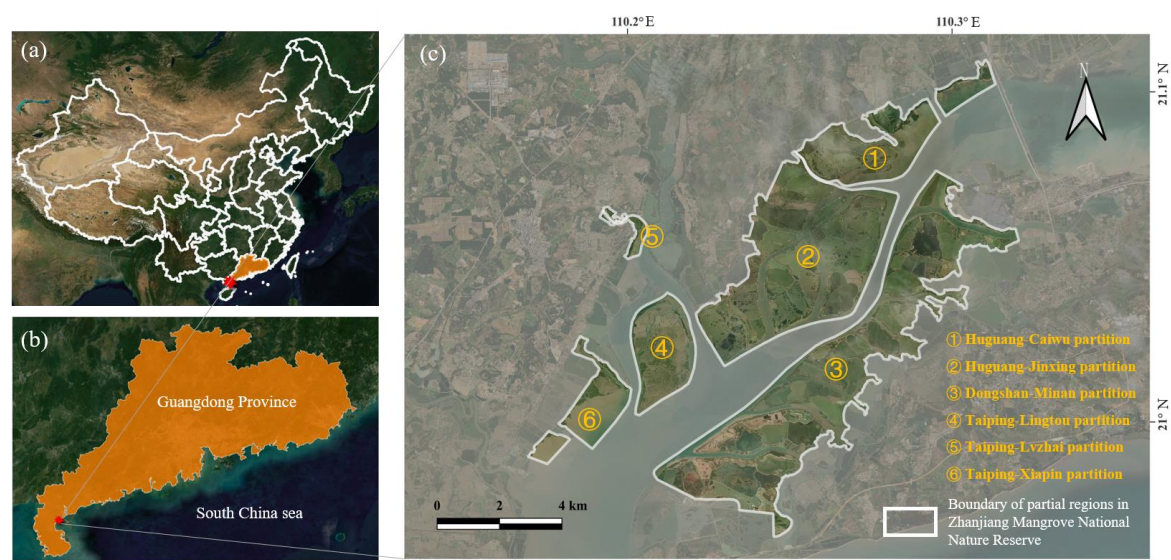


Figure 1. (a) The location of mangrove study area in China. The orange area is the Guangdong province, and the red pentagram represents the study area; (b) zoom in the study area; (c) the partial scope and brief info of Zhanjiang mangrove national nature reserve where the study area is located.

2.2. Imagery Data from DJI drones

This study employed a DJI Phantom 4RTK multispectral drone to undertake the initial data collection phase in October 2023 for the mangrove area in Jinniu island. The drone is equipped with an all-in-one multispectral imaging system that integrates a visible camera and five multispectral cameras[47], including blue, green, red, red-edge, and near-infrared (see **Error! Reference source not found.** for details of the spectral ranges of each band). This system is capable of fulfilling the tasks of both visible and multispectral imaging[48]. This drone is equipped with an RTK (Real-Time Kinematic, real-time Dynamic Carrier Phase Difference technology) positioning module, which offers a position accuracy of 1 cm ± 1 ppm horizontally and 1.5 cm ± 1 ppm vertically[49]. One ppm represents an error increase of 1 mm for every 1 km the aircraft travels[49]. Consequently, this equipment can meet the required position accuracy standards without establishing ground control points during in-situ operations.

Table 1. Spectral range and resolution of the Phantom 4RTK multispectral drone sensor.

Spectral name	Spectral range and tolerance (nm)	Spatial resolution (In this study) (cm)
Blue band (B)	450 ± 16	4.61
Green band (G)	560 ± 16	
Red band (R)	650 ± 16	
Red Edge band (RE)	730 ± 16	
Near Infrared (NIR)	840 ± 16	
Visible light band (RGB)	390 ~ 780	

A total of 3726 multispectral images were acquired in this data acquisition work, with the flight altitude at 100 m, the side overlap degree of 70%, and the forward overlap degree of 80%. The elevation angle of the sensor was set at -90 degrees, the photo mode was equal time interval, the flight speed was 10.2 m/s, and the ground sample distance (GSD) was 4.61 cm/pixel in ortho mode. The resolution of the drone image is considerably higher than that of the commercial satellite image, which has the highest resolution (0.15 m) and significant potential for identifying and classifying the spatial distribution of specific mangrove species. These drone imageries show that differential mangrove species exhibit distinctive characteristics. The *Rhizophora stylosa* association is presented by a nearly circular dark green feature in the image. At the same time, the *Aegiceras corniculatum*-*Avicennia marina* association is also depicted as a green feature, albeit with a slightly lighter hue than

that of the *Rhizophora stylosa* association. Additionally, the area of the *Aegiceras corniculatum*-*Avicennia marina* association is relatively limited, exhibiting a scattered distribution. In the image, the *Sonneratia apetala* association presents a lighter green feature, and the edge features are more conspicuous (Error! Reference source not found.).

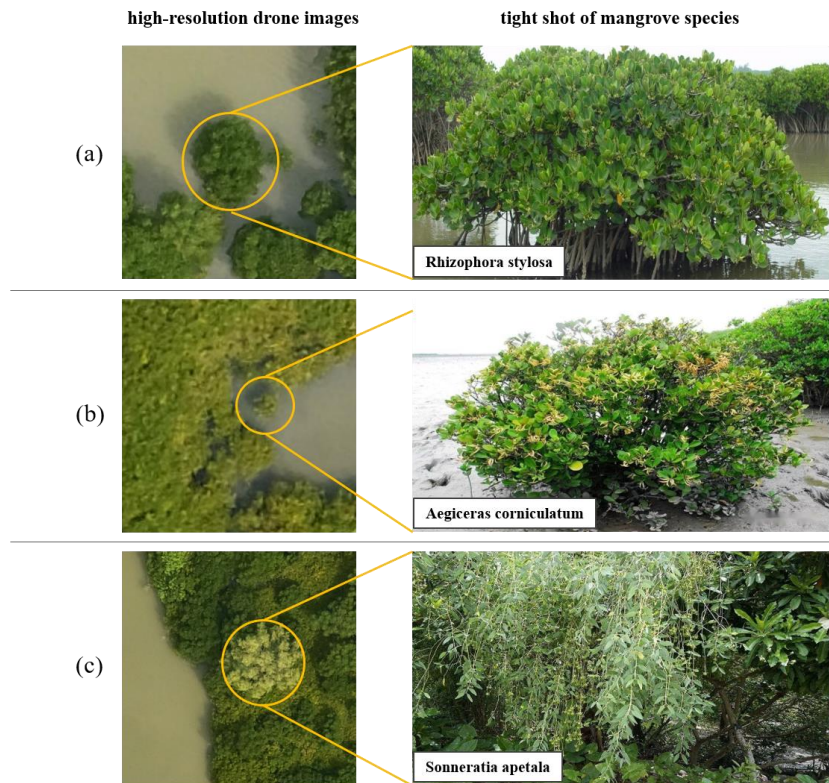


Figure 2. Illustrates the distinctive performance of different mangrove types in high-resolution drone images. (a) *Rhizophora stylosa* in high-resolution drone image (left) and tight shot of zooming in image (right); (b) and (c) similar image cases of *Aegiceras corniculatum* and *Sonneratia apetala*.

3. Methodology

3.1. Deep Neural network Model Design

3.1.1. Application of Attention Mechanisms in Deep Learning

The attention mechanism[50] was integrated into the basic model structure to enhance the capability of extracting mangrove areas. The attention mechanism is a simple and effective module for convolutional neural networks dealing with 2D image data[50]. For the feature layers generated during the convolution process, the attention module obtained feature layers with channel-distributed weights and spatial-distributed weights from the channel and spatial dimensions, respectively. These two feature layers with weight information were then multiplied by the corresponding initial feature layers to complete the attention to the channel or spatial features. This attention mechanism can be integrated into the training process of the convolutional neural network as a generalized lightweight module so that the channel and spatial information in the image can be fully exploited. Among them, the computational process of the channel attention module is as follows:

$$W_c(F_{prev}) = \delta \left(MLP \left(AvgPool(F_{prev}) \right) MLP \left(MaxPool(F_{prev}) \right) \right) \quad (1)$$

Where F_{prev} is the result of the encoding process of the neural network model; MLP is a multilayer perceptron with shared weights; $AvgPool$ is the average pooling operation; and $MaxPool$ is the maximum pooling operation.

As the input data of the channel attention module, the size of the result F_{prev} of the coding process was $H \times W \times C$. Then, global average pooling and global max pooling were performed to

obtain two semantic feature layers, the size of which was $1 \times 1 \times C$. These features were then fed into a weighted multi-layer perceptron, and *ReLU* was used as the activation function to enhance the non-linear relationship between the feature layers. The *Sigmoid* activation function was used to obtain the weight coefficients of the different channels. Finally, a new feature layer based on channel attention could be obtained by multiplying the obtained channel weight coefficient with the original input feature layer.

The processing of the spatial attention mechanism is similar to the channel attention mechanism for an input feature layer of size $H \times W \times C$. First, average pooling and max pooling were performed along the channel direction to obtain a feature layer with the same number of layers and channels with spatial feature information[50]. Then, the results of average pooling and maximum pooling were concatenated. Then, a 7×7 convolution kernel was used for the convolution calculation, and the spatial feature weight coefficients were obtained after processing with a sigmoid activation function. Finally, the original input feature layer was multiplied by the spatial feature weight coefficients to get the new feature layer based on the spatial attention mechanism.

$$W_s(F_{\text{prev}}) = \delta(f^{n \times n}[\text{AvgPool}(F_{\text{prev}}); \text{MaxPool}(F_{\text{prev}})]) \quad (2)$$

Where $f^{n \times n}$ is an $n \times n$ convolution process. Finally, the new feature layer based on the channel attention mechanism and the new feature layer based on the spatial attention mechanism were linked and input to the subsequent decoder. As a result, the deep learning model can fully use the band information and spatial feature information of the remote sensing image to more accurately identify and extract the mangrove areas and species[51].

3.1.2. Novel Model of Integrating Multi-Scale Spectral and Spatial Information for Detecting Mangrove Area

For effectively extracting mangrove area, we designed MangroveNet, a novel enhanced model with integrating multi-scale attentional visual field information. The overall model structure is shown in **Error! Reference source not found.**. The model comprises three main components: an encoding structure, a corresponding decoding mechanism, and a multi-scale attention extraction module. In terms of sampling mode, the model imitates the sampling mode of SegNet architecture[52]. In order to enhance the model more lightweight, particularly in scenarios where hardware memory is limited, only two downsamples of maximum pooling operations were implemented. Consequently, this results in the formation of three distinct scale-specific feature layer views. It is postulated that during the processes of convolution and semantic information extraction, there exist critical spectral and spatial information that necessitates consideration across diverse scales of vision. The attention mechanism module can then be utilized to assign significant weights to these focused spectral or spatial information. Subsequently, the feature layer containing the high-weight spectral information and the feature layer comprising the spatial information are connected, thereby yielding a feature layer that integrates the essential information from both spectral and spatial perspectives. In this paper, the focus is on the critical spectral and spatial characteristics of the mangrove area. The integrated spectral and spatial information feature layer ($X_{\text{con_att}}^l$) can be obtained according to the following calculation formula:

$$X_{\text{con_att}}^l = W_c^l \times X^l + W_s^l \times X^l \quad (3)$$

Where W_c^l is the weights from channel attention; W_s^l is weights from spatial attention; l represents the number of feature layers; X^l represents feature map in l -th feature layer. Subsequently, the feature layer must be concatenated with the convolution feature layer that has not been subjected to maximum pooling. However, given that the size of the feature layer is equivalent to that prior to the attention mechanism module, it is the feature layer after the maximum pooling. Consequently, it is necessary to upsample it prior to its connection with the convolution feature layer that has not been sampled by max pooling. It is important to note that if the feature layer has been maximized twice, the number of up-samples should also be doubled to ensure that the initial size can be restored. The feature information layer with weight information from each scale is fused with the initial feature layer and we can obtain $X_{\text{con_total}}$:

$$X_{con_total} = \sum_{i=1}^n (UpSamp(X_{con_att}^i) + X_{con_total}^{i-1}) \quad (4)$$

Where n indicates the number of times about max pooling operations; $X_{con_total}^{i-1}$ represents the result of the last integration. Specifically, when $i-1=0$, $X_{con_total}^0$ represents the initial layer have been convolved but not max-pooled; $UpSamp$ is the upsampling operations. In addition, the operations for low-resolution feature layers entail the retention of index information from each encoder maximum pooling operation, which is then employed in upsampling scheme at the decoding stage. After integrating high-weight spectral and spatial information of all scales and normal convolution sampling information, the feature layers were fed into *Sigmoid* for segmentation[52], the formulation is indicated as follows:

$$X_{sigmoid} = \delta \left(\frac{1}{1 + e^{-(X_{con_total} + X_{upsampling}^n)}} \right) \quad (5)$$

Where n is the number of times about upsampling operations; $X_{upsampling}^n$ represents the feature map after the n -th upsampling. In the training process, it is also necessary to compare the predicted result with the label value by using the cross-entropy loss function, and send the error back to the network for adjustment, until the predicted result meets the threshold for output. The loss function ($L_{mangrove}$) can be calculated using the following formulation:

$$L_{mangrove} = -\frac{1}{n} \sum_{i=1}^n (X_i^t \log(X_i^p) + (1 - X_i^t) \log(1 - X_i^p)) \quad (6)$$

Where n is the total pixel number of features; X_i^t represents the true value and X_i^p is the prediction value.

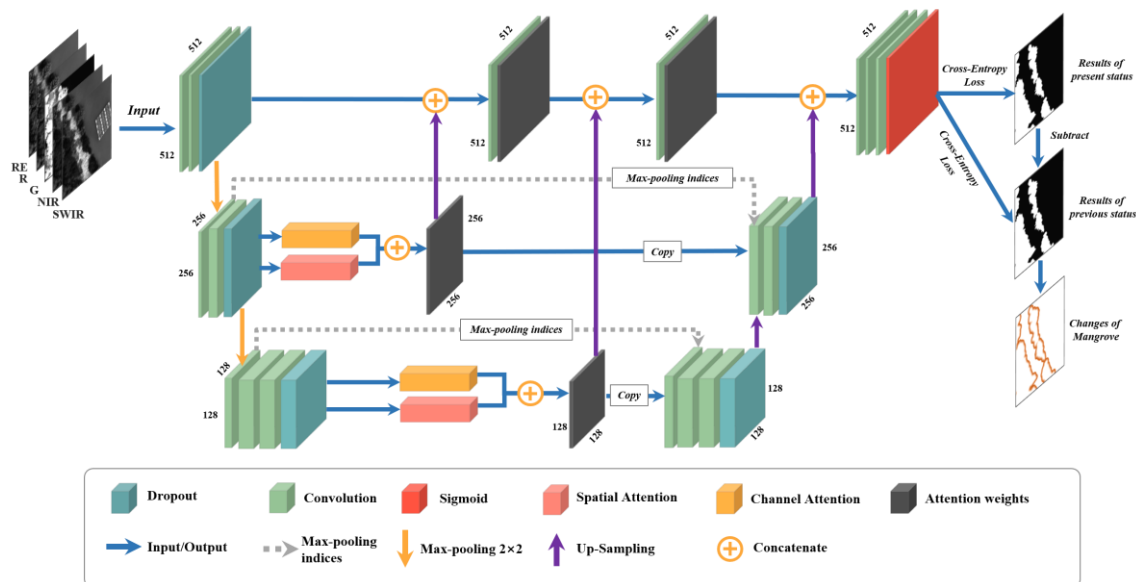


Figure 3. The internal structure of the MangroveNet model.

3.1.3. Improved Model for Identifying Mangrove Species

Due to mangrove species identification is more complex than mangrove extraction, we choose a more complex model—CloudNet+ as the base model for species identification. CloudNet+ is a fully convolutional neural network model with a deep learning encoder-decoder structure[54], mainly used for cloud detection and segmentation in Landsat-8 optical imagery. The model has several encapsulated blocks, including Feedforward Block, Expanding Block, Contracting Block, and Upsampling Block, as shown in **Error! Reference source not found.** The Contracting Block contains

three successive convolutional layers and a parallel 1×1 convolutional layer, and after both complete the convolutional operation, the results of each are summed and then maximally pooled[55]. The Feedforward Block joined and pooled the results obtained from the Contracting Block, and finally, the results of each branch were pooled and accumulated[55]. The input data of the Expanding Block came from the Contracting Block and the Feedforward Block. The Feedforward Block result was convolved twice, summed with the Contracting Block, and then fed into the Upsampling Block. The result of the Feedforward Block was convolved twice and summed with the Contracting Block. Finally, the Upsampling Block performed an upsampling operation on the incoming feature layer to restore the image to its original size. The backward propagation of these blocks was trained using the Filtered Jaccard Loss function, which avoided the misjudgment of predicted values caused by the absence of valid values in the labels. Following the success of CloudNet+ in segmenting cloud-covered areas in remote sensing images, this study builds on its foundations and improves it to identify mangrove species. Given the complexity of the model, we have elected to augment the channels and spatial attention mechanisms between down sampling and upsampling. This approach is intended to take full advantage of the spectral and spatial information of the images, thereby forming an improved AttCloudNet+ model.

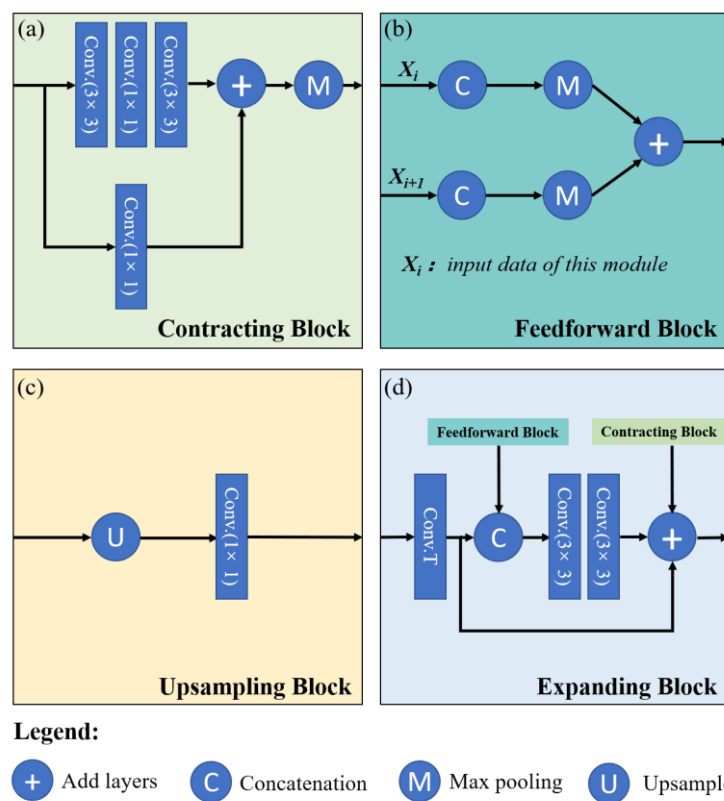


Figure 4. Schematic diagram of the internal structure in the base model[55]. (a) Contracting Block; (b) Feedforward Block; (c) Upsampling Block; (d) Expanding Block.

3.2. Data Pre-Processing

The limitations of the field of view and the high degree of overlap in drone photography mean that the small aerial drone images taken during the mission must first be stitched together. Due to the inevitable occurrence of some distortions during shooting, the overlapping regions of the images will produce a certain degree of bias during stitching, so it is necessary to eliminate some severe distortions or excessively unsmooth regions after stitching. In this study, the Agisoft Metashape 2.1.1 software[56] (<https://www.agisoft.com/>, Agisoft Metashape allows the processing of images from visible or multispectral cameras into dense point clouds, textured polygon models, georeferenced true orthophotos and spatial information data in the form of Digital Surface Models, accessed on 28 May 2024) was used to process the acquired high-resolution multispectral drone images into a dense

point cloud, textured polygon models, DSMs and georeferenced true orthophotos. The acquired high-resolution multispectral drone images were mosaicked, and orthophotos of the study area were generated. The drone remote sensing images of the mangrove areas in each band were mosaicked with an average resolution of approximately 13137×15061 pixels. A total of 18 multi-phase sampling orthophotos were generated and stored in TIFF format.

Drone images are often affected by the shape of the earth's surface and map projections during acquisition and transmission. Therefore, geographic correction is required to ensure the accuracy and reliability of the geographic information. Geo-correction is a geometric correction method that realizes the geo-localization of images by establishing an accurate correlation[57] between the pixels of a drone remote sensing image and the geographic coordinate points on the earth's surface. This correction corrects for image deformation due to the earth's curvature, map projection, and surface undulation. The principle of geo-correction in this study includes the following steps:

(1) Selection of control points: Before geo-correction, a set of control points must be selected. In this study, the accuracy of the geo-correction process was ensured by uniformly selecting five different geographic feature points within the image of the mangrove area.

(2) Performing the correction: During the correction process, each image pixel was moved to an exact position in the geographic coordinate system through geometric transformations (e.g., polynomial transformation or cubic spline interpolation) based on the known control points.

In this study, the 16-bit multispectral drone orthoimages acquired after the above operations were converted to 8-bit orthoimages[58] using the Geospatial Data Abstraction Library (GDAL, <https://www.gdal.org>, accessed on 15 July 2024). Then, the geographic coordinate system in all images was converted to the Universal Transverse Mercator (UTM) projection coordinate system using GDAL. Within the mangrove area of Zhanjiang, the reference of the projected coordinate system of the image was WGS_1984_UTM_Zone_49N. Since the original image size is large for the deep learning model, each multispectral drone orthoimage was segmented into an image of 512×512 pixels to create the image training dataset for the deep neural network in this study. During the segmentation process, an overlap of 150 pixels was ensured between each image and its neighboring images, which ensured both a sufficient number of training datasets and utilized the edge feature information of the images effectively.

3.3. Identify Mangrove Areas and Species

Error! Reference source not found. shows the flowchart of this study. In this study, MangroveNet was employed to extract the extent of mangrove areas from high-resolution multispectral drone imagery. AttCloudNet+ was employed to identify the spatial distribution of different species within the mangrove area. Multispectral images of mangrove forests and their surrounding areas on Jinniu Island were selected as training and validation samples for this study. During the prediction step, the two models were applied to their respective corresponding data, namely high-resolution multispectral drone images of mangrove forests in the mangrove study area on Jinniu Island.

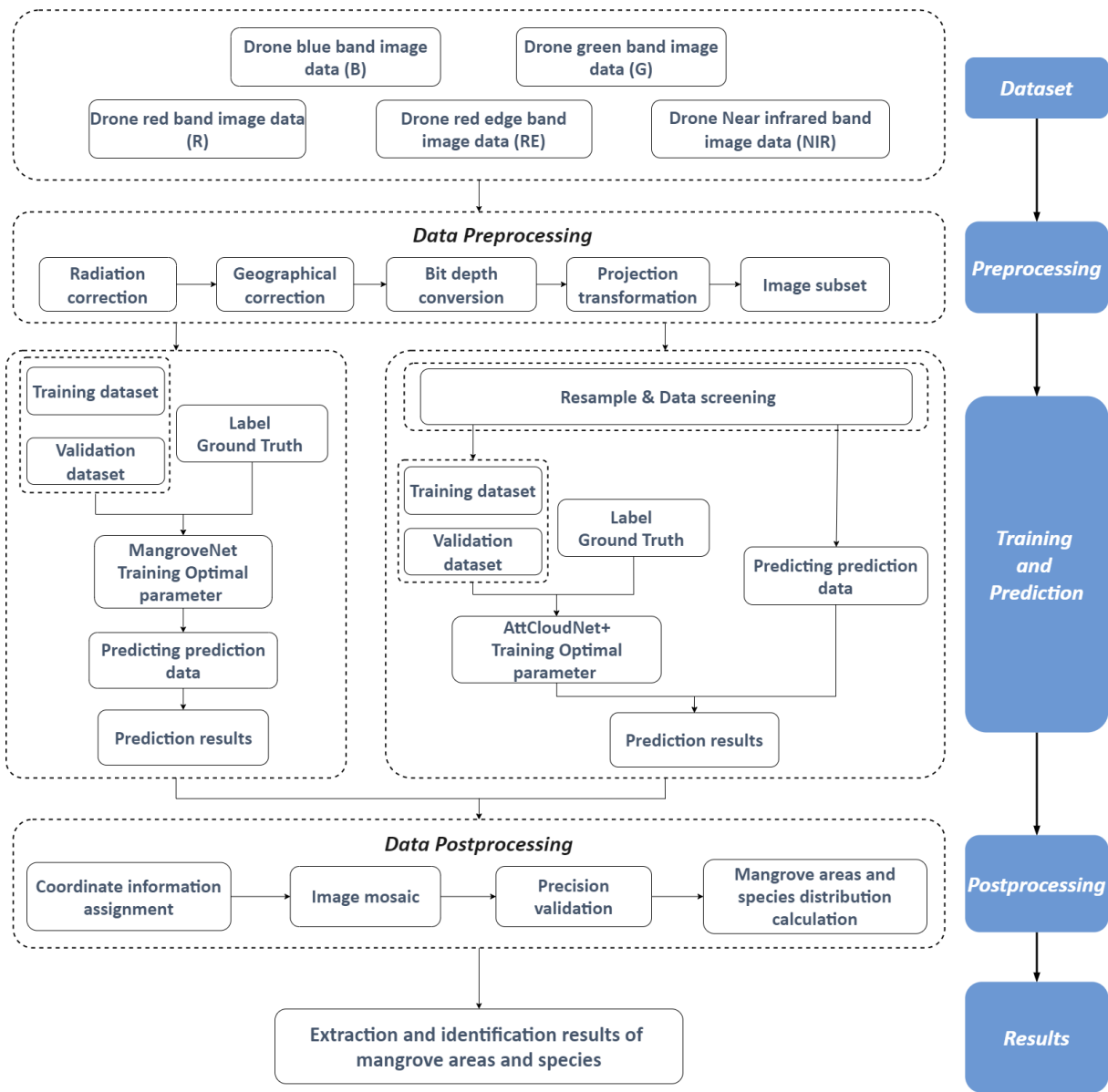


Figure 5. Flowchart of mangrove extraction and species identification using drone imagery.

3.3.1. Training Dataset Generation

The multispectral drone imagery training data used in this study was collected from the DJI Phantom 4RTK multispectral sensor. The area chosen for generating the training dataset is located in and around the mangrove forest of Jinniu Island. This location was selected because the mangrove forests on Jinniu Island are lush and well-protected, making the boundary between the water and the mangrove areas more distinct. Nevertheless, we selected different methods for mangrove range extraction and mangrove species identification. For mangrove range extraction, 25 subranges were selected as training data in this study, as shown in **Error! Reference source not found.** (a).

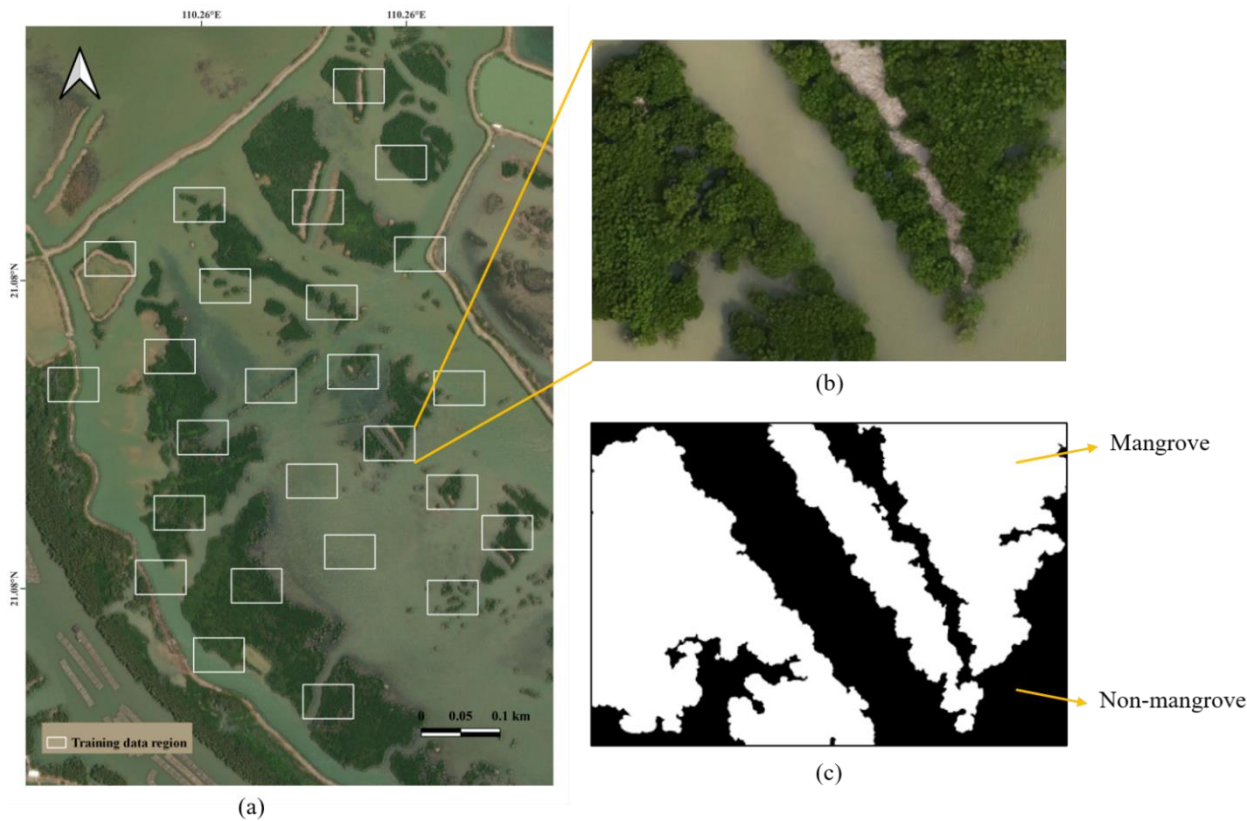


Figure 6. (a) Training dataset region for mangrove range extraction; (b) Example of a specifically selected training range; (c) Label of the corresponding sample training range.

An integrated mangrove region with concentrated and abundant mangrove species in this area was selected for species identification. The training dataset for the mangrove species identification task was also selected from the Jinniu Island mangrove forest, focusing on areas with a high concentration of mangrove species types. The total area of this region is 3.12 hectares. The field investigation revealed that the mangrove tree species prevalent in this region include *Rhizophora stylosa*, *Aegiceras corniculatum*, *Sonneratia apetala*, and *Avicennia marina*. In some areas, there is a notable association between *Aegiceras corniculatum* and *Avicennia marina*, manifesting as *Aegiceras corniculatum*-*Avicennia marina* tree clusters (**Error! Reference source not found.**).

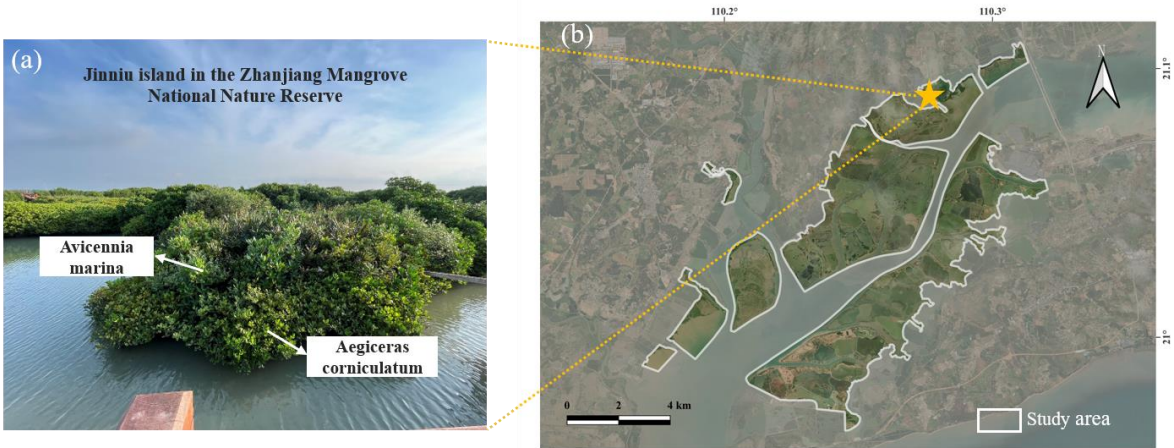


Figure 7. (a) In-situ photos of associated growth of *Aegiceras corniculatum* and *Avicennia marina*; (b) scope of the study area and the yellow pentagram represents the sampling position.

Accordingly, the primary mangrove species identifications in this integrated mangrove region are the *Rhizophora stylosa* association, the *Aegiceras corniculatum*-*Avicennia marina* association,

the *Sonneratia apetala* association, and the non-mangrove ground objects (e.g., water bodies, aquaculture fishing rows, and roads) as shown in **Error! Reference source not found.**

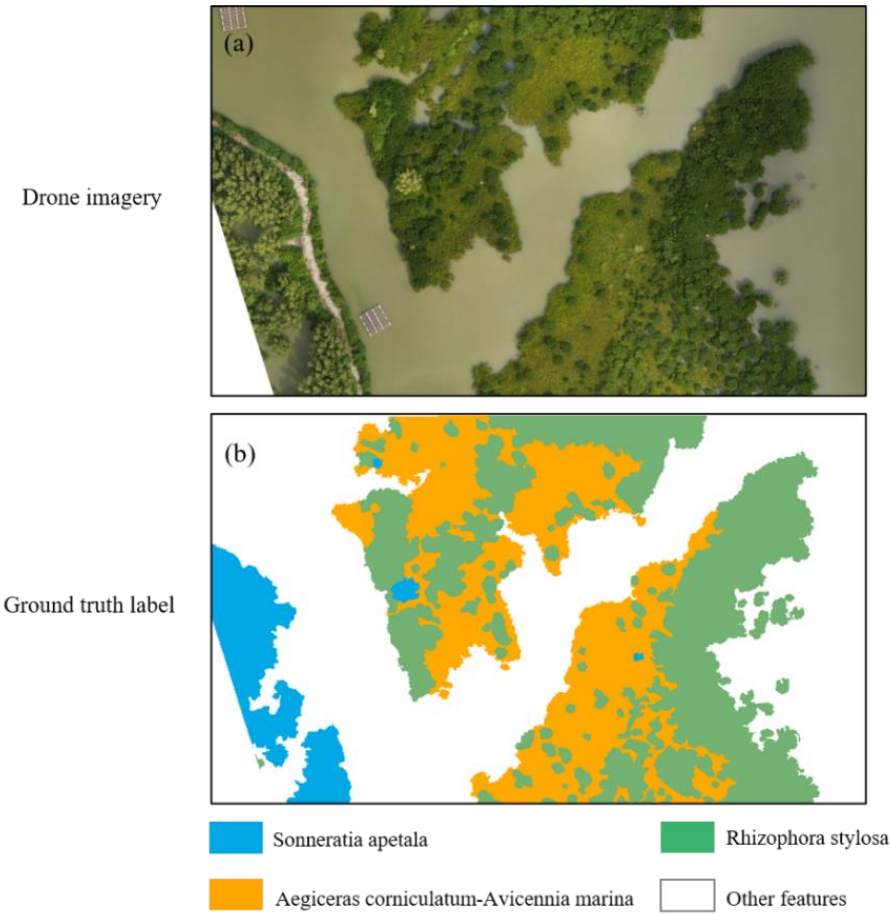


Figure 8. Training dataset region for mangrove species identification. (a) drone in-situ image of the selected mangrove region; (b) the corresponding mangrove species type label for this specific region.

To generate the label dataset[59], we first tried the commonly used annotation tool Labelme (graphical interface annotation software widely used for image annotation)[60]. However, it was found that the edges of the labels labeled by Labelme were relatively smooth, which could not meet the experimental requirements of accurate segmentation of the mangrove area and its species categories. Therefore, in this study, different labeling strategies were used for the two parts of the research tasks concerning the characteristics of each band of the drone multispectral image. For the mangrove area extraction task, this study used a threshold extraction combined with a visual interpretation method to accurately annotate the multispectral images to achieve high accuracy, as shown in **Error! Reference source not found.** (b–c). First, the mangrove area was roughly identified using the exponential thresholding method, and then the misidentified areas were corrected by experienced personnel. It was challenging to perform automated labeling for the mangrove species identification task due to the number of classification categories. Therefore, this study adopted a direct manual labeling method for the mangrove species identification task and used expert experience to ensure the accuracy of labeling different mangrove species, as shown in **Error! Reference source not found.** As the area of the label was also the same as the area of the image, each label also had to be segmented and cropped into 512 × 512 pixels. In the segmentation process, each image had to overlap 150 pixels with the neighboring images, as in the image segmentation process, to ensure that the label matched the image data.

3.3.2. Training and Prediction

In this part, we first describe our experimental setting. The experimental platform of this study employs an Intel 11th Gen Intel® Core™ i7-11700@2.50GHz 2.50GHz processor equipped with 32.0G memory and a NVIDIA GeForce GTX 4060Ti graphics card. In the experiment, Anaconda3 (64-bit) was employed as the carrier for environment configuration. The Windows 10 Professional 64-bit operating system was utilized to create The Keras tool, which is integrated into the framework and is used to construct the model[61]. The CUDA 11.2.2 version corresponds to the computer configuration and is selected as the GPU computing platform, and cuDNN 8.9.2 is installed as the deep learning GPU acceleration library. Finally, PyCharm Community Edition 2021.3.3 is employed to develop and compile the deep learning program.

The training dataset was prepared using high-resolution multispectral drone remote sensing images in five bands: B, G, R, RE, and NIR. In the mangrove range extraction task, 29,808 segmented training data were obtained in each band after pre-processing. 23,848 segmented images (80% of all segmented images) were used to train MangroveNet, and the remaining 5,960 sub-images were used as a validation dataset in the training process[62]. The mangrove species distribution prediction task obtained 10560 segmented training data for each band. Among them, 8448 segmented images (80% of all segmented images) were used to train AttCloudNet+[62], and the remaining 2112 sub-images were used as a validation dataset in the training process. In the training process, the learning rate was set to 0.0001, the number of training in the same batch (Batch size) was set to 4, and the maximum number of iterations (Maximun iteration) was set to 1000. After training this network, this study predicted the extent of mangrove forests and the spatial distribution status of their species in the prediction data collected from the mangrove area of Jinniu island, in which the prediction data contained a total of 3,120 segmented sub-images.

3.4. Accuracy Metrics

3.4.1. Evaluation Criteria of Mangrove Range Extraction

In this study, five evaluation indicators were selected to assess the prediction results of each model, including Accuracy, F1-Score, Precision, Recall, and mIoU[63] (Mean Intersection over Union), as shown in equations (7) to (11).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

$$mIoU = \frac{1}{2} \times \left(\frac{TP}{TP+FP+FN} + \frac{TN}{TN+FN+FP} \right) \quad (11)$$

TP , TN , FP , and FN correspond to true positives, true negatives, false positives, and false negatives, respectively, in the confusion matrix of the prediction results[64], as shown in **Error! Reference source not found.**(a). TP is the correctly categorized mangrove pixels; TN is the correctly

categorized non-mangrove pixels; *FP* is the incorrectly categorized non-mangrove pixels incorrectly identified as mangrove; and *FN* is the mangrove pixels incorrectly identified as non-mangrove. The confusion matrix could then be used to construct a variety of metrics to evaluate the results. *Accuracy* represents the proportion of correctly recognized pixels to the total number of pixels; *Precision* is defined as the proportion of correctly recognized mangrove pixels to the number of pixels recognized as mangrove; *Recall* represents the proportion of correctly recognized mangrove pixels to the number of real mangrove pixels in the labels; *F1-Score* reveals the weights summed average of *Precision* and *Recall*; and *mIoU* measures the degree of similarity of the model's prediction results to the real labels, as shown in **Error! Reference source not found.** (b).

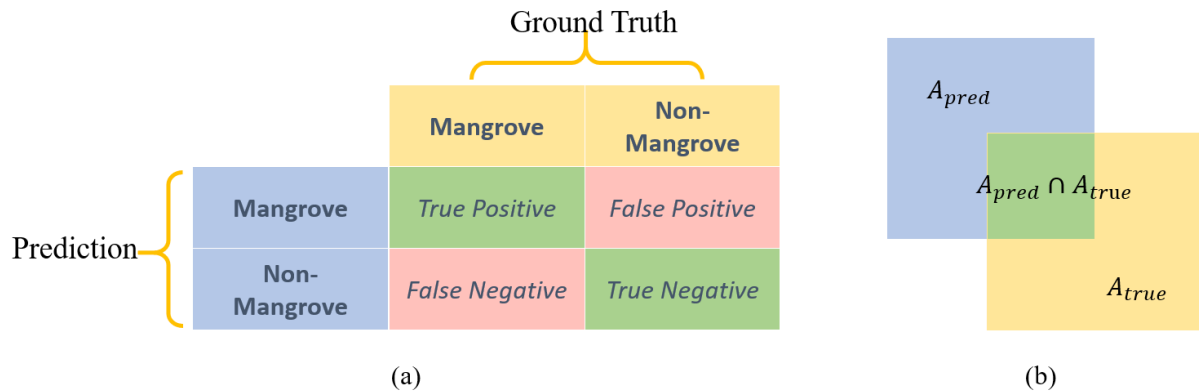


Figure 9. (a) Schematic of confusion matrix; (b) schematic of *mIoU* calculation process.

3.4.2. Indicators of Mangrove Species Identification

Due to the considerable number of identification results in the mangrove species identification task, various evaluation strategies are employed to assess the accuracy of the results. The selected criteria for evaluating the identification accuracy of AttCloudNet+ and other traditional supervised classification, unsupervised classification, common deep learning algorithms, and machine learning algorithms are the Confusion Matrix, Kappa coefficient, and Overall Accuracy[65].

The Confusion Matrix, also known as the Error Matrix, is an array used to compare the number of image elements classified into a category and the number of ground tests for that category. Typically, the matrix columns represent the reference data, and the rows represent the category data obtained from the classification results of the remote sensing data. The Kappa coefficient is an index that can show more comprehensively the degree of consistency between the predicted results and the true labels (**Error! Reference source not found.**). The overall accuracy can indicate the general accuracy of the classification results without considering the situation of a specific category. The Kappa coefficient can solve this problem well and analyze the classification accuracy more comprehensively and objectively. The expression of the Kappa coefficient is as follows:

$$Kappa = \frac{Accuracy - P_e}{1 - P_e} \quad (12)$$

$$P_e = \frac{(TP+FN) \times (FP+TN) \times (TP+FP) \times (FN+TN)}{(TP+FP+FN+TN)^2} \quad (13)$$

where *Accuracy* represents the observed agreement probability; P_e is the expected agreement probability, which represents the expected observed agreement probability in the case of randomness. The explanation of *TP*, *TN*, *FP*, and *FN* shown in **Error! Reference source not found.**(a).

Table 2. Kappa factor evaluation criteria[66].

Range of coefficient values	Evaluation criteria
Kappa = -1	Completely inconsistent
Kappa = 0	Random classification results
0 < Kappa ≤0.2	Notably low consistency
0.2 < Kappa ≤0.4	Low consistency
0.4 < Kappa ≤0.6	Medium consistency
0.6 < Kappa ≤0.8	High consistency
0.8 < Kappa <1.0	Almost complete consistency
Kappa = +1	Complete consistency

The Overall Accuracy (OA) equals the sum of correctly classified pixels divided by the total number of pixels. The number of correctly classified pixels is distributed along the diagonal of the confusion matrix, and the total number of pixels is equal to the total number of pixels from all actual reference sources. The overall classification accuracy is the probability that, for each random sample, the classified result matches the test data type.

4. Results

In consideration of the spatial and multispectral characteristics inherent to high-resolution drone remote sensing imagery, MangroveNet and AttCloudNet+ were employed to detect the extent of mangrove forests and to identify the spatial distribution of mangrove species. A comparison was conducted between the initial results obtained from our models and other mainstream semantic segmentation models, including UNet, SegNet, and other UNet-based variant neural network models. Moreover, results under other traditional supervised or unsupervised algorithms were also compared with ours. The quantitative comparison results of the area of interest in the mangrove reserve region are demonstrated in

and **Error! Reference source not found..**

4.1. Mangrove Area Extraction and Evaluation

Table 3 reveals the performance of other diverse segmentation models and our promoted model on the evaluation criteria.

Table 3 is based on five metrics – Accuracy, F1-score, mIoU, Precision, and Recall. We obtained the following conclusions from

: (1) The MangroveNet model exhibits optimal performance in Accuracy, F1-Score, mIoU, and Precision. Regarding the recall indicator, AttCloudNet+ demonstrated the most favorable performance. (2) The distinctive spatial and spectral characteristics of ground objects in remote sensing images present a challenge for the conventional deep learning model based on convolution and pooling layers, which cannot adequately attend to and enhance mangroves' spatial and spectral characteristics in remote sensing images. Consequently, models other than MangroveNet perform poorly in most indicators. (3) Similarly, although AttCloudNet+ is equipped with spatial and spectral attention mechanisms, its considerable model size hinders its performance in mangrove-non-mangrove binary classification. The subsequent mangrove species segmentation experiments also demonstrated that AttCloudNet+ is better suited to mangrove species' multi-category identification tasks. Furthermore, we focused on comparing the fluctuations in Loss and mIoU across a range of deep learning models, including MangroveNet, AttCloudNet+, DeepUNet, and SegNet, about the duration of the training process. Following over 35 training rounds for all models, both Loss and mIoU demonstrated a stable state. Compared to other models, MangroveNet exhibited the lowest Loss value, while the mIoU index demonstrated the highest rate of increase and reached the highest value after 20 training rounds.

Table 3. Performance of different deep learning models in mangrove area extraction.

Model Structure	Accuracy	F1_Score	mIoU	Precision	Recall
UNet	94.52%	96.20%	94.52%	94.52%	99.95%
DeepUNet	99.13%	92.10%	91.04%	93.15%	91.89%
ResUNet	97.13%	97.47%	96.00%	96.00%	99.95%
SegNet	96.71%	97.20%	95.63%	95.64%	99.95%
AttCloudNet+	95.18%	96.50%	94.97%	94.97%	99.96%
MangroveNet	99.13%	98.84%	98.11%	99.62%	98.38%

Each model's results were visualized to facilitate comparison between the models and ascertain the effect of each model (**Error! Reference source not found.**). A comparative analysis showed that the prediction results of MangroveNet were the most closely aligned with the actual in-situ circumstances. Conversely, the prediction results of alternative models exhibited varying degrees of misidentification. From the results of each model after visualization, we can compare and find: (1) While creating the labels, the water surface within the mangrove gap was not identified. However, it is important to note that the mangrove community is not a single entity; it is distributed with water surfaces in its gap. The UNet, DeepUNet, ResUNet, and SegNet models could not accurately identify the water surface within the mangrove gap. In contrast, the MangroveNet and AttCloudNet+ demonstrated superior performance, with MangroveNet offering a more precise reflection of the actual ground situation. This outcome suggests incorporating spatial and spectral channel attention mechanisms can enhance recognition accuracy. (2) The UNet model erroneously identified a bare road as a mangrove area, while the AttResUNet only partially excluded the aquaculture fisheries, erroneously identifying them as mangroves. Furthermore, the edge of the mangrove forest was identified with considerable ambiguity in the recognition results of ResUNet. Based on the recognition outcomes of these models, ResUNet exhibited the least effective recognition performance. MangroveNet represents the optimal model, offering a comprehensive and intuitive approach to performance evaluation and a robust quantitative assessment of indicators.

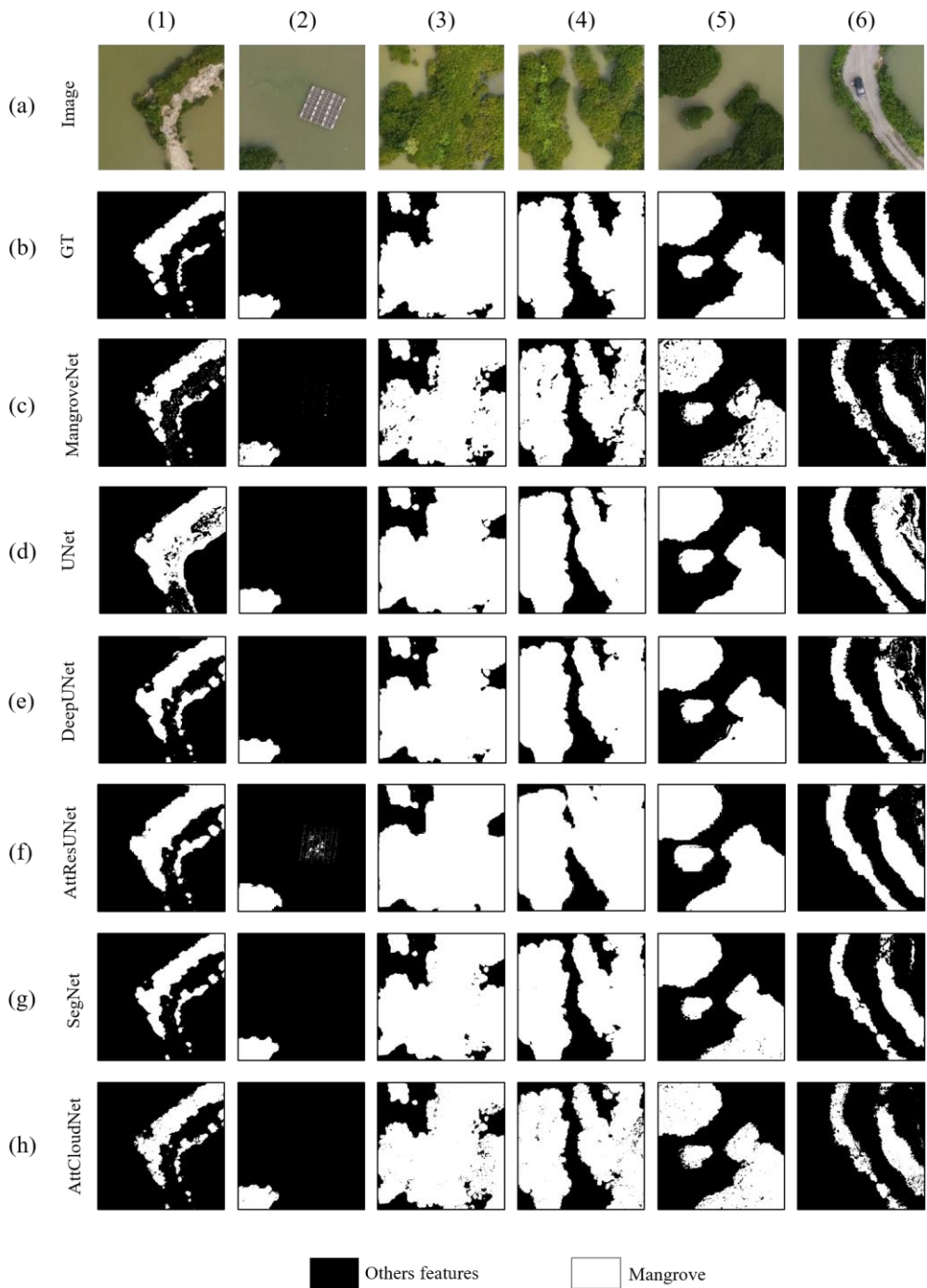


Figure 10. Comparison of different models' prediction results of mangrove area. Row (a) represents drone images; row (b–h) extraction results of different models. Columns (1–6) are diverse sites of drone images.

4.2. Mangrove Species Identification and Evaluation

Furthermore, the predictions of AttCloudNet+ are presented (**Error! Reference source not found.**) and compared with those of other traditional supervised and unsupervised models, machine learning models, and deep learning models (**Error! Reference source not found.**). The results demonstrate that the AttCloudNet+ prediction is the most accurate, while SegNet exhibits superior performance in predicting the maturation of *Sonneratia apetala* but inferior performance in predicting *Aegiceras corniculatum* and *Rhizophora stylosa*. The SVM demonstrates a robust prediction efficacy for the *Rhizophora stylosa* and *Aegiceras corniculatum*, yet exhibits a comparatively diminished prediction efficacy for the *Sonneratia apetala*. Additionally, RF demonstrated a notable predictive capability for *Rhizophora stylosa* and *Aegiceras corniculatum*. However, it is noteworthy that this model exhibited a certain degree of inaccuracy in identifying the

Sonneratia apetala, and its capacity to eliminate the influence of aquaculture fishing rows needed to be improved. K-means can identify the range of the *Aegiceras corniculatum* to a certain extent; however, it cannot accurately identify the range of the other two mangroves and cannot eliminate the influence of aquaculture fishing rows. ISODATA misidentified the *Rhizophora stylosa* as the *Sonneratia apetala*, the *Aegiceras corniculatum* as the *Rhizophora stylosa*, and the *Sonneratia apetala* as the *Aegiceras corniculatum*, and was unable to eliminate the influence of aquaculture fishing rows.

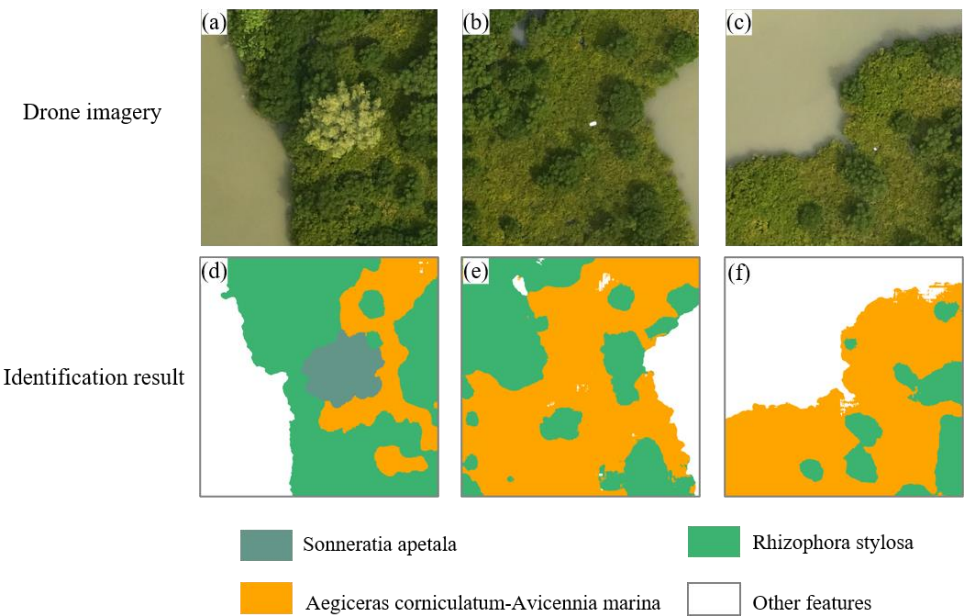


Figure 11. (a—c) Mangrove species images captured by drone; (d—f) identification results of mangrove species spatial distribution using AttCloudNet+.

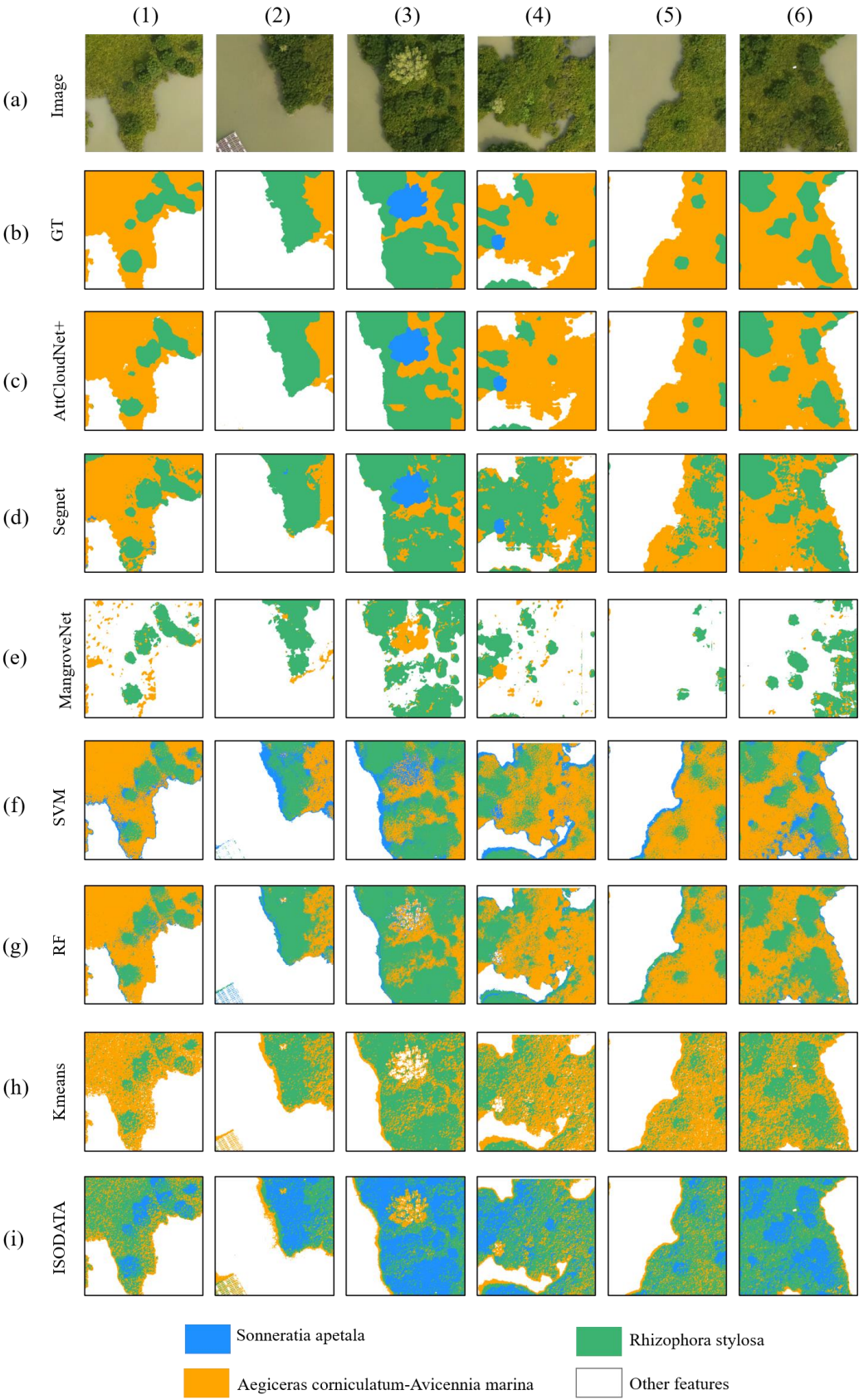


Figure 12. Comparison of different models' prediction results of mangrove species. Row (a) represents drone images; row (b–i) mangrove species identification results using different deep learning and conventional methods. Columns (1–6) are diverse sites of drone images.

We calculated the confusion matrix of mangrove major species categories. The confusion matrix of species identification results (**Error! Reference source not found.**) in this study demonstrated that

following the implementation of the enhanced AttCloudNet+ for predicting potential species within drone images, the identification accuracy of the diverse group prediction results is as follows: *Aegiceras corniculatum*-*Avicennia marina* (86%), *Rhizophora stylosa* (86%), *Sonneratia apetala* (72%), and other objects (90%). The primary cause of the misidentification of *Rhizophora stylosa* is the erroneous classification of 11% of *Aegiceras corniculatum*-*Avicennia marina* as *Rhizophora stylosa*. This may be attributed to the fact that the color and spectral characteristics of *Aegiceras corniculatum*-*Avicennia marina* and *Rhizophora stylosa* exhibit a similar spatial distribution, which may result in the misidentification of the model. The model's prediction of *Sonneratia apetala* exhibited the lowest accuracy, primarily due to the misidentification of 19% of *Sonneratia apetala* as *Rhizophora stylosa*. This discrepancy may be attributed to the limited availability of *Sonneratia apetala* samples in the sampling area, which impairs the model's ability to discern the characteristics of *Sonneratia apetala* during training effectively. Additionally, the similarity in morphology between *Sonneratia apetala* and other nearby mangrove species may contribute to the confusion.

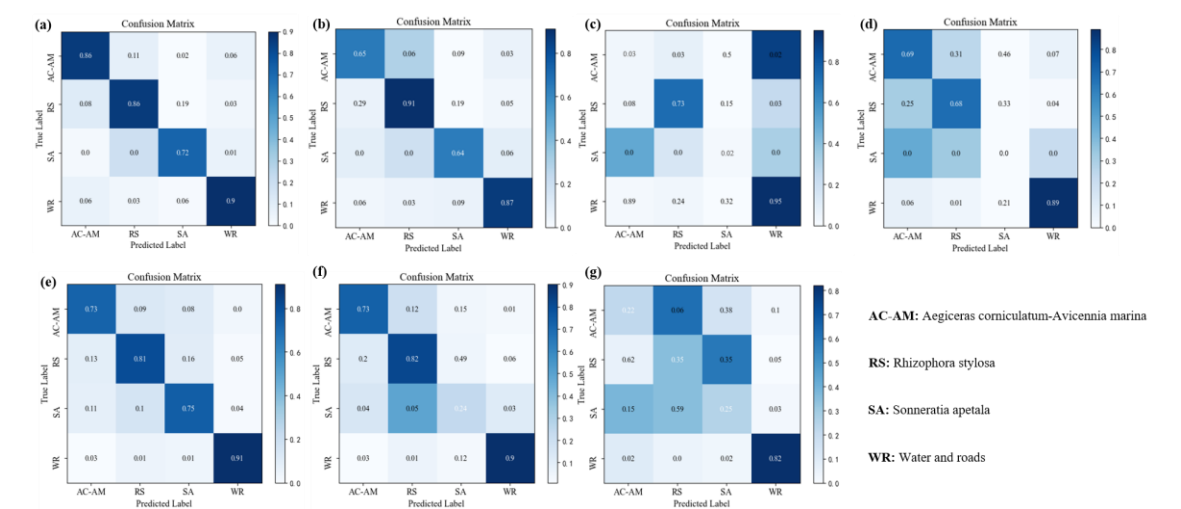


Figure 13. The confusion matrix of species identification results under different methods. The horizontal axis of each graph represents the category of the predicted results, and the vertical axis represents the category of the true value. (a) AttCloudNet+; (b) SegNet; (c) MangroveNet; (d) K-means; (e) SVM; (f) Random forest; (g) ISODATA.

presents a comparative analysis of the AttCloudNet+ with traditional supervised classification, unsupervised classification, machine learning, and other methodologies, including the K-means clustering approach, ISODATA cluster analysis, and the Random Forest (RF) model. The performance of the Kappa coefficient and the overall accuracy (OA) index of the mangrove species identification by RF, Support Vector Machine (SVM), and other methods was evaluated.

demonstrated that the mangrove species identification prediction results using AttCloudNet+ exhibited the optimal performance in terms of the Kappa coefficient and the OA index. ISODATA has the worst Kappa coefficient and OA performance because the difference between mangrove species could be more apparent. For ISODATA, similar categories are often merged, leading to classification results that differ significantly from reality. Compared to traditional supervised or unsupervised classification algorithms, machine learning and deep learning algorithms show high Kappa coefficient and OA results, showing that machine learning and deep learning algorithms have great advantages in multi-category segmentation tasks of remote sensing images.

Table 4. Precision comparison of mangrove species identification results by different methods.

Methods	Kappa coefficient	Overall Accuracy (OA)
K-means	0.61	0.75
ISODATA	0.36	0.56

Random Forest	0.71	0.81
SVM	0.76	0.84
SegNet	0.73	0.82
MangroveNet	0.41	0.67
AttCloudNet+	0.81	0.87

4.3. Calculation of Distribution Area

This study calculated the area of mangroves and their species using the area range deep learning model predicted.

i. To calculate the mangrove range, the grid mosaic calculation of the mangrove range prediction results is first performed. This is followed by the use of the TabulateArea function in the ArcPy Python API (ArcPy is a Python library for ArcGIS Desktop and ArcGIS Pro), which is employed to determine the distribution of pixels classified as mangroves within the specified area. Subsequently, the mangrove range area S is statistically calculated according to the pixel area size (Equation 14), where i is the number of pixels, r is the ground resolution represented by pixels, and N is the total number of pixels within the target range.

$$S = \sum_{i=1}^N r_i^2$$

(14)

ii. About the distribution area of the mangrove species, it is evident that the mangrove species included in the predicted results have their respective result values (**Error! Reference source not found.**). Thus, pixel statistics for each mangrove category must be conducted using the values resulting from the above. By calculating the number of pixels in a specific area and the area of a single pixel, it is possible to obtain the spatial distribution area of each mangrove species within the area. The mangrove species range area S_j was calculated statistically (Equation 15), where i is the number of pixels corresponding with the mangrove species, j is the species' category represented by the grey value of the different prediction results, r is the ground resolution represented by pixels, and N is the total number of pixels within the target range.

$$S_j = \sum_{i=1}^N r_i^2$$

(15)

Once the spatial distribution of mangrove groups had been obtained, a detailed area distribution calculation was carried out to explore the specific quantitative parameters of mangrove group distribution. This was done for various group distribution conditions. In the prediction process, the gray values of the feature markers representing various group categories were assigned the following labels: *Aegiceras corniculatum*-*Avicennia marina* (we set the gray value to 0), *Rhizophora stylosa* (gray value = 1), *Sonneratia apetala* (gray value = 2), and other features (gray value = 3). Accordingly, to ascertain the distribution area of the various mangrove groups in the prediction results, the number of grey values corresponding to each group can be subjected to statistical analysis by the partition. This will enable the distribution area of the various groups in the final recognition result to be determined according to the pixel size. Due to the lack of published mangrove species distribution data set in this region and to verify the accuracy of the mangrove species identification model for the distribution area of various groups, this study manually identified and calculated the distribution area of the groups above within the study area. The resulting calculation and comparison are presented in **Error! Reference source not found.**. The accuracy of species areas (A_{sa}) calculated by the model was defined by the areas of various mangrove groups calculated by the model and the areas of various mangrove groups depicted artificially. The A_{PA} was calculated statistically (Equation 16), where i is a serial number representing a mangrove species, S_{Di} is the area calculated by manual delineation (**Error! Reference source not found.**), S_{Mi} is the area calculated by AttCloudNet+ (**Error! Reference source not found.**), and N is the total number of mangrove species. The resulting value

from this calculation was 87.66%, indicating an average accuracy of the mangrove species area calculation.

$$A_{sa} = \frac{1}{N} \sum_{i=1}^N \frac{|S_{D_i} - S_{M_i}|}{S_{D_i}} \tag{16}$$

Table 5. Results comparison of mangrove species identification model and artificial recognition.

Mangrove species	AttCloudNet+ (m²)	Manual delineation (m²)
Aegiceras corniculatum-Avicennia marina	6873.33	6013.33
Rhizophora stylosa	7960.00	7680.00
Sonneratia apetala	1386.67	1713.33

5. Discussion of Issues

5.1. Factors Affecting the Results of Each Model

In the extraction of mangrove areas, this study showed low F1-score, Precision, Recall, and mIoU metrics when using other models (e.g., UNet, DeepUNet, ResUNet, CloudNet+, and SegNet) for the mangrove area identification extraction, as shown in

. These models showed low efficiency, which may be affected by the following reasons: UNet had the lowest metric score. Due to the simple structure of the model, it may not be able to handle the complex texture features of high-resolution drone images. DeepUNet is a variant of UNet that considers the image's spectral features. However, the method ignored the spatial feature information, leading to uncertainty in identifying mangrove areas. SegNet has similar limitations to DeepUNet. Since ResUNet only considers spatial features and ignores spectral features, it may not perform as well as MangroveNet.

5.2. Impact of Shadows on Model Predictions

While extracting mangrove area identification, it was found that the results of mangrove area extraction were not satisfactory in some areas. This is because the area was affected by the shadow of the mangrove projections in the light, resulting in these two types of features having close grey values and texture features in the image. This caused the model to incorrectly recognize the shadow of the mangrove forest in the light as part of the mangrove area in the prediction process, making the actual prediction result of the mangrove forest large.

This is because the model used in this study was temporarily unable to distinguish mangrove light shadows from normal mangrove areas. In future studies, mangrove image data with light shadows can be tried as negative samples to correct the model's misrecognition of shaded areas in the training data. In addition, linear stretching of the raw multispectral drone image data can mitigate the effect of mangrove light shadows on the prediction process.

6. Conclusions

In this study, we proposed enhanced deep learning models, namely MangroveNet and AttCloudNet+, to facilitate the detection of changes in the extent of mangrove habitats in the study area. This was achieved by identifying and extracting the salient features present in high-resolution multispectral drone spatial imagery obtained from the mangrove area of Jinniu Island in Zhanjiang Mangrove National Nature Reserve and the spatial distribution of mangrove species within the aforementioned study area. The accuracy validation results show that the MangroveNet model used in this study has a high degree of accuracy in identifying and extracting mangrove areas. Compared to other mainstream models, the accuracy of MangroveNet in the prediction process was improved

to 99.13%, while the *mIoU* was improved to 98.11%. Moreover, our MangroveNet model shows the most optimal performance across the metrics: *Accuracy*, *mIoU*, *Precision*, and *F1_Score*.

The confusion matrix revealed that the classification accuracy of the various species prediction results reached 86% for *Aegiceras corniculatum*-*Avicennia marina*, 86% for *Rhizophora stylosa*, 72% for *Sonneratia apetala*, and 90% for other features, respectively, following the application of AttCloudNet+ to predict the possible mangrove species categories in the image. Moreover, a comparison was conducted between the Kappa coefficient and overall accuracy (OA) of AttCloudNet+ and those of traditional supervised classification, unsupervised classification, and machine learning methods for species category prediction. The results demonstrate that the Kappa coefficient of AttCloudNet+ is 0.81, indicating that the model's predictions for mangrove species are highly consistent with the actual situation. Furthermore, the overall classification accuracy of AttCloudNet+ reaches 0.87, which is the optimal performance among the six traditional supervised classification, unsupervised classification, machine learning, and other methods.

In conclusion, the MangroveNet and AttCloudNet+ proposed in this paper can efficiently and accurately obtain mangrove ranges and detect changes based on high-resolution multispectral drone spatial image data. Moreover, they can accurately predict and analyze the spatial distribution status of mangrove species. This approach offers an enhanced spatial and temporal resolution of mangrove forests and their species, providing a robust foundation for evaluating mangrove protection, monitoring, and assessing tropical and subtropical forest distribution. It offers scientific and technological support for assessing and monitoring mangrove protection effectiveness.

Author Contributions: X.C.: conceptualization, methodology, algorithm, formal analysis, data curation, visualization, writing—original draft. X.Z.: writing—review and editing, investigation, funding acquisition. C.Z.: formal analysis, writing—review and editing, supervision, project administration. X.D.: formal analysis and writing—original draft. L.K., Z.X., and X.H.: validation and visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Innovation Project of the Guangdong Academy of Environmental Sciences (No. XMLX-20230707017) and the Special Fund Project for Environmental Protection of Guangdong Province (No.4, 2024).

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study.

Acknowledgments: We acknowledge the Zhanjiang Mangrove National Nature Reserve Bureau for supporting the field data collection in this research project.

Conflicts of Interest: The authors state that they have no competing financial interests or personal relationships which may have influenced the work presented in this paper.

References

1. Friess, D.A.; Rogers, K.; Lovelock, C.E.; Krauss, K.W.; Hamilton, S.E.; Lee, S.Y.; Lucas, R.; Primavera, J.; Rajkaran, A.; Shi, S. The State of the World's Mangrove Forests: Past, Present, and Future. *Annual Review of Environment and Resources* **2019**, *44*, 89-115, doi:https://doi.org/10.1146/annurev-environ-101718-033302.
2. Akram, H.; Hussain, S.; Mazumdar, P.; Chua, K.O.; Butt, T.E.; Harikrishna, J.A. Mangrove Health: A Review of Functions, Threats, and Challenges Associated with Mangrove Management Practices. *Forests* **2023**, *14*, 1698.
3. Schubert, M.; Machkour-M'Rabet, S.; Santamaría-Rivero, W.; Dzib-Chay, A.; Torres Cristiani, L.; MacKinnon-Haskins, B. Multi-character approach reveals a new mangrove species of the Yellow Warbler complex, *Setophaga petechia*, on Cozumel Island, Mexico. *Plos One* **2023**, *18*, doi:10.1371/journal.pone.0287425.
4. Jia, M.; Wang, Z.; Mao, D.; Ren, C.; Song, K.; Zhao, C.; Wang, C.; Xiao, X.; Wang, Y. Mapping global distribution of mangrove forests at 10-m resolution. *Science Bulletin* **2023**, *68*, 1306-1316, doi:10.1016/j.scib.2023.05.004.

5. Tran, T.V.; Reef, R.; Zhu, X.; Gunn, A. Characterising the distribution of mangroves along the southern coast of Vietnam using multi-spectral indices and a deep learning model. *Science of The Total Environment* **2024**, *923*, 171367, doi:<https://doi.org/10.1016/j.scitotenv.2024.171367>.
6. Kuenzer, C.; Bluemel, A.; Gebhardt, S.; Quoc, T.V.; Dech, S. Remote Sensing of Mangrove Ecosystems: A Review. *Remote Sens.* **2011**, *3*, 878-928. <https://doi.org/10.3390/rs3050878>
7. Seydi, S.T.; Ahmadi, S.A.; Ghorbanian, A.; Amani, M. Land Cover Mapping in a Mangrove Ecosystem Using Hybrid Selective Kernel-Based Convolutional Neural Networks and Multi-Temporal Sentinel-2 Imagery. *Remote Sens.* **2024**, *16*, 2849.
8. Dey, B.; Ahmed, R.; Ferdous, J.; Haque, M.M.U.; Khatun, R.; Hasan, F.E.; Uddin, S.N. Automated plant species identification from the stomata images using deep neural network: A study of selected mangrove and freshwater swamp forest tree species of Bangladesh. *Ecological Informatics* **2023**, *75*, 102128, doi:<https://doi.org/10.1016/j.ecoinf.2023.102128>.
9. Adhiwibawa, M.A.; Ariyanto, M.R.; Struck, A.; Prilianti, K.R.; Brotosudarmo, T.H. Convolutional neural network in image analysis for determination of mangrove species. In Proceedings of the Third International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2018), **2019**; pp. 13-20.
10. Arias-Aguilar, J.A.; López-Jimenez, E.; Ramírez-Cárdenas, O.D.; Herrera-Lozada, J.C.; Hevia-Montiel, N. Mangrove semantic segmentation on aerial images. *IEEE Latin America Transactions* **2024**, *22*, 379-386, doi:10.1109/TLA.2024.10500718.
11. Markham, B.L.; Storey, J.C.; Williams, D.L.; Irons, J.R. Landsat sensor performance: history and current status. *IEEE transactions on geoscience and remote sensing* **2004**, *42*, 2691-2694.
12. Kpienbaareh, D.; Sun, X.; Wang, J.; Luginaah, I.; Bezner Kerr, R.; Lupafya, E.; Dakishoni, L. Crop Type and Land Cover Mapping in Northern Malawi Using the Integration of Sentinel-1, Sentinel-2, and PlanetScope Satellite Data. *Remote Sens.* **2021**, *13*, 700. <https://doi.org/10.3390/rs13040700>
13. Gerace, A.D.; Schott, J.R.; Nevins, R. Increased potential to monitor water quality in the near-shore environment with Landsat's next-generation satellite. *Journal of Applied Remote Sensing* **2013**, *7*, 073558.
14. Xu, M.; Jia, X.; Pickering, M.; Jia, S. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *149*, 215-225, doi:<https://doi.org/10.1016/j.isprsjprs.2019.01.025>.
15. Kleinschroth, F.; Banda, K.; Zimba, H.; Dondeyne, S.; Nyambe, I.; Spratley, S.; Winton, R.S. Drone imagery to create a common understanding of landscapes. *Landscape and Urban Planning* **2022**, *228*, 104571, doi:<https://doi.org/10.1016/j.landurbplan.2022.104571>.
16. Gray, P.C.; Ridge, J.T.; Poulin, S.K.; Seymour, A.C.; Schwantes, A.M.; Swenson, J.J.; Johnston, D.W. Integrating Drone Imagery into High Resolution Satellite Remote Sensing Assessments of Estuarine Environments. *Remote Sens.* **2018**, *10*, 1257.
17. Curran, P. Multispectral remote sensing of vegetation amount. *Progress in Physical Geography: Earth and Environment* **1980**, *4*, 315-341, doi:10.1177/030913338000400301.
18. Gxokwe, S.; Dube, T.; Mazvimavi, D. Multispectral Remote Sensing of Wetlands in Semi-Arid and Arid Areas: A Review on Applications, Challenges and Possible Future Research Directions. *Remote Sens.* **2020**, *12*, 4190.
19. Thakur, S.; Mondal, I.; Ghosh, P.B.; Das, P.; De, T.K. A review of the application of multispectral remote sensing in the study of mangrove ecosystems with special emphasis on image processing techniques. *Spatial Information Research* **2020**, *28*, 39-51, doi:10.1007/s41324-019-00268-y.
20. Zimudzi, E.; Sanders, I.; Rollings, N.; Omlin, C.W. Remote sensing of mangroves using unmanned aerial vehicles: current state and future directions. *Journal of Spatial Science* **2021**, *66*, 195-212, doi:10.1080/14498596.2019.1627252.
21. Ali, A.; Nayyar, Z.A. Extraction of mangrove forest through Landsat 8 Mangrove Index (L8MI). *Arabian Journal of Geosciences* **2020**, *13*, 1132, doi:10.1007/s12517-020-06138-4.
22. Baloloy, A.B.; Blanco, A.C.; Sta. Ana, R.R.C.; Nadaoka, K. Development and application of a new mangrove vegetation index (MVI) for rapid and accurate mangrove mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *166*, 95-117, doi:<https://doi.org/10.1016/j.isprsjprs.2020.06.001>.

23. Yang, G.; Huang, K.; Sun, W.; Meng, X.; Mao, D.; Ge, Y. Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. *ISPRS Journal of Photogrammetry and Remote Sensing* **2022**, *189*, 236-254, doi:<https://doi.org/10.1016/j.isprsjprs.2022.05.003>.
24. Shi, T.; Liu, J.; Hu, Z.; Liu, H.; Wang, J.; Wu, G. New spectral metrics for mangrove forest identification. *Remote Sensing Letters* **2016**, *7*, 885-894, doi:10.1080/2150704X.2016.1195935.
25. Diniz, C.; Cortinhas, L.; Nerino, G.; Rodrigues, J.; Sadeck, L.; Adami, M.; Souza-Filho, P.W.M. Brazilian Mangrove Status: Three Decades of Satellite Data Analysis. *Remote Sens.* **2019**, *11*, doi:10.3390/rs11070808.
26. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* **2002**, *83*, 195-213, doi:[https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).
27. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sensing of Environment* **1994**, *48*, 119-126, doi:[https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1).
28. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **2007**, *17*, 1425-1432, doi:10.1080/01431169608948714.
29. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9*, 62-66, doi:10.1109/TSMC.1979.4310076.
30. Ji, L.; Zhang, L.; Wylie, B. Analysis of dynamic thresholds for the normalized difference water index. *Photogrammetric Engineering and Remote Sensing* **2009**, *75*, 1307-1317, doi:10.14358/PERS.75.11.1307.
31. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255-260, doi:10.1126/science.aaa8415.
32. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *114*, 24-31, doi:<https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
33. Yue, S.; Li, P.; Hao, P. SVM classification: Its contents and challenges. *Applied Mathematics-A Journal of Chinese Universities* **2003**, *18*, 332-342, doi:10.1007/s11766-003-0059-5.
34. Basheer, I.A.; Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods* **2000**, *43*, 3-31, doi:[https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3).
35. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS journal of photogrammetry and remote sensing* **2018**, *145*, 60-77.
36. Xu, C.; Wang, J.; Sang, Y.; Li, K.; Liu, J.; Yang, G. An Effective Deep Learning Model for Monitoring Mangroves: A Case Study of the Indus Delta. *Remote Sens.* **2023**, *15*, 2220.
37. Dang, K.B.; Nguyen, M.H.; Nguyen, D.A.; Phan, T.T.H.; Giang, T.L.; Pham, H.H.; Nguyen, T.N.; Tran, T.T.V.; Bui, D.T. Coastal Wetland Classification with Deep U-Net Convolutional Networks and Sentinel-2 Imagery: A Case Study at the Tien Yen Estuary of Vietnam. *Remote Sens.* **2020**, *12*, 3270.
38. Faza, S.; Nababan, E.B.; Efendi, S.; Basyuni, M.; Rahmat, R.F. An initial study of deep learning for mangrove classification. *IOP Conference Series: Materials Science and Engineering* **2018**, *420*, 012093, doi:10.1088/1757-899X/420/1/012093.
39. Krishna, K.; Murty, M.N. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **1999**, *29*, 433-439, doi:10.1109/3477.764879.
40. Leempoel, K.; Satyanarayana, B.; Bourgeois, C.; Zhang, J.; Chen, M.; Wang, J.; Bogaert, J.; Dahdouh-Guebas, F. Dynamics in mangroves assessed by high-resolution and multi-temporal satellite data: a case study in Zhanjiang Mangrove National Nature Reserve (ZMNNR), P. R. China. *Biogeosciences* **2013**, *10*, 5681-5689, doi:10.5194/bg-10-5681-2013.
41. Zhou, X.; Chen, M.; Liang, C. Optimal schemes of groundwater exploitation for prevention of seawater intrusion in the Leizhou Peninsula in southern China. *Environmental Geology* **2003**, *43*, 978-985, doi:10.1007/s00254-002-0722-9.
42. Zheng, Z.; Lei, Z.Q. A 400,000 year record of vegetational and climatic changes from a volcanic basin, Leizhou Peninsula, southern China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **1999**, *145*, 339-362, doi:[https://doi.org/10.1016/S0031-0182\(98\)00107-2](https://doi.org/10.1016/S0031-0182(98)00107-2).

43. Xue, J.; Zhong, W.; Xie, L.; Unkel, I. Vegetation responses to the last glacial and early Holocene environmental changes in the northern Leizhou Peninsula, south China. *Quaternary Research* **2015**, *84*, 223-231, doi:10.1016/j.yqres.2015.08.001.
44. Wang, G.; Zhang, Y.; Guan, D.; Xiao, L.; Singh, M. The potential of mature *Sonneratia apetala* plantations to enhance carbon stocks in the Zhanjiang Mangrove National Nature Reserve. *Ecological Indicators* **2021**, *133*, 108415, doi:https://doi.org/10.1016/j.ecolind.2021.108415.
45. Lu, C.; Li, L.; Wang, Z.; Su, Y.; Su, Y.; Huang, Y.; Jia, M.; Mao, D. The national nature reserves in China: Are they effective in conserving mangroves? *Ecological Indicators* **2022**, *142*, 109265, doi:https://doi.org/10.1016/j.ecolind.2022.109265.
46. Durango-Cordero, J.; Satyanarayana, B.; Chan, J.C.-W.; Bogaert, J.; Dahdouh-Guebas, F. Distinguishing Original and Non-Original Stands at the Zhanjiang Mangrove National Nature Reserve (P.R. China): Remote Sensing and GIS for Conservation and Ecological Research. *Remote Sens.* **2021**, *13*, 2781.
47. Taddia, Y.; Stecchi, F.; Pellegrinelli, A. Coastal Mapping Using DJI Phantom 4 RTK in Post-Processing Kinematic Mode. *Drones* **2020**, *4*, 9.
48. Wang, N.; Guo, Y.; Wei, X.; Zhou, M.; Wang, H.; Bai, Y. UAV-based remote sensing using visible and multispectral indices for the estimation of vegetation cover in an oasis of a desert. *Ecological Indicators* **2022**, *141*, 109155, doi:https://doi.org/10.1016/j.ecolind.2022.109155.
49. Peppas, M.V.; Hall, J.; Goodyear, J.; Mills, J.P. PHOTOGRAMMETRIC ASSESSMENT AND COMPARISON OF DJI PHANTOM 4 PRO AND PHANTOM 4 RTK SMALL UNMANNED AIRCRAFT SYSTEMS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *XLII-2/W13*, 503-509, doi:10.5194/isprs-archives-XLII-2-W13-503-2019.
50. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), **2018**; pp. 3-19.
51. Guo, Y.; Liao, J.; Shen, G. Mapping Large-Scale Mangroves along the Maritime Silk Road from 1990 to 2015 Using a Novel Deep Learning Model and Landsat Data. *Remote Sens.* **2021**, *13*, 245.
52. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, *39*, 2481-2495, doi:10.1109/TPAMI.2016.2644615.
53. Theckedath, D.; Sedamkar, R.R. Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN Computer Science* **2020**, *1*, 79, doi:10.1007/s42979-020-0114-9.
54. Khandelwal, P.; Armstrong, S.; Matin, A.; Pallickara, S.; Pallickara, S.L. CloudNet: A Deep Learning Approach for Mitigating Occlusions in Landsat-8 Imagery using Data Coalescence. In *Proceedings of the 2022 IEEE 18th International Conference on e-Science (e-Science)*, 11-14 Oct. 2022, **2022**; pp. 117-127.
55. Mohajerani, S.; Saeedi, P. Cloud-Net+: A Cloud Segmentation CNN for Landsat 8 Remote Sensing Imagery Optimized with Filtered Jaccard Loss Function. *ArXiv* **2020**, abs/2001.08768.
56. Over, J.-S.R.; Ritchie, A.C.; Kranenburg, C.J.; Brown, J.A.; Buscombe, D.D.; Noble, T.; Sherwood, C.R.; Warrick, J.A.; Wernette, P.A. Processing coastal imagery with Agisoft Metashape Professional Edition, version 1.6—Structure from motion workflow documentation; 2021-1039; Reston, VA, **2021**; p. 46.
57. Goncalves, H.; Goncalves, J.A.; Corte-Real, L. Measures for an Objective Evaluation of the Geometric Correction Process Quality. *IEEE Geoscience and Remote Sensing Letters* **2009**, *6*, 292-296, doi:10.1109/LGRS.2008.2012441.
58. Başeski, E. 16-Bit to 8-Bit Conversion in Remote Sensing Images by Using Image Content. In *Proceedings of the 2019 9th International Conference on Recent Advances in Space Technologies (RAST)*, 11-14 June 2019, **2019**; pp. 413-417.
59. Bernhardt, M.; Castro, D.C.; Tanno, R.; Schwaighofer, A.; Tezcan, K.C.; Monteiro, M.; Bannur, S.; Lungren, M.P.; Nori, A.; Glocker, B.; et al. Active label cleaning for improved dataset quality under resource constraints. *Nature Communications* **2022**, *13*, 1161, doi:10.1038/s41467-022-28818-3.
60. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision* **2008**, *77*, 157-173, doi:10.1007/s11263-007-0090-8.
61. Chicho, B.T.; Bibi Sallow, A. A Comprehensive Survey of Deep Learning Models Based on Keras Framework. *Journal of Soft Computing and Data Mining* **2021**, *2*, 49-62.

62. Barbedo, J.G.A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture* **2018**, *153*, 46-53, doi:<https://doi.org/10.1016/j.compag.2018.08.013>.
63. Liu, L.; Wu, Y.; Wei, W.; Cao, W.; Sahin, S.; Zhang, Q. Benchmarking Deep Learning Frameworks: Design Considerations, Metrics and Beyond. In *Proceedings of the 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2-6 July 2018, **2018**; pp. 1258-1269.
64. Tien Bui, D.; Hoang, N.-D.; Martínez-Álvarez, F.; Ngo, P.-T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of The Total Environment* **2020**, *701*, 134413, doi:<https://doi.org/10.1016/j.scitotenv.2019.134413>.
65. Foody, G.M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment* **2020**, *239*, 111630, doi:<https://doi.org/10.1016/j.rse.2019.111630>.
66. Thompson, W.D.; Walter, S.D. A REAPPRAISAL OF THE KAPPA COEFFICIENT. *Journal of Clinical Epidemiology* **1988**, *41*, 949-958, doi:[https://doi.org/10.1016/0895-4356\(88\)90031-5](https://doi.org/10.1016/0895-4356(88)90031-5).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.