remote sensing

MDPI

Article

Semantic Segmentation of China's Coastal Wetlands Based on Sentinel-2 and Segformer

Xufeng Lin [1], Youwei Cheng [1], Gong Chen [1], Wenjing Chen [1], Rong Chen [2], Demin Gao [1], Yinlong Zhang [2,3,4] and Yongbo Wu [2,3,*]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; linxufeng@njfu.edu.cn (X.L.); youwei@njfu.edu.cn (Y.C.); gongchen@njfu.edu.cn (G.C.); wjchen@njfu.edu.cn (W.C.); dmgao@njfu.edu.cn (D.G.)
[2] College of Ecology and Environment, Nanjing Forestry University, Nanjing 210037, China; rongchen@njfu.edu.cn (R.C.); ylzhang@njfu.edu.cn (Y.Z.)
[3] Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
[4] National Positioning Observation Station of Hung-Tse Lake Wetland Ecosystem in Jiangsu Province, Huaian 223100, China
* Correspondence: yongbowu@njfu.edu.cn; Tel.: +86-025-85428935

Abstract: Concerning the ever-changing wetland environment, the efficient extraction of wetland information holds great significance for the research and management of wetland ecosystems. China's vast coastal wetlands possess rich and diverse geographical features. This study employs the SegFormer model and Sentinel-2 data to conduct a wetland classification study for coastal wetlands in Yancheng, Jiangsu, China. After preprocessing the Sentinel data, nine classification objects (construction land, *Spartina alterniflora (S. alterniflora)*, *Suaeda salsa (S. salsa)*, *Phragmites australis (P. australis)*, farmland, river system, aquaculture and tidal falt) were identified based on the previous literature and remote sensing images. Moreover, mAcc, mIoU, aAcc, Precision, Recall and F-1 score were chosen as evaluation indicators. This study explores the potential and effectiveness of multiple methods, including data image processing, machine learning and deep learning. The results indicate that SegFormer is the best model for wetland classification, efficiently and accurately extracting small-scale features. With mIoU (0.81), mAcc (0.87), aAcc (0.94), mPrecision (0.901), mRecall (0.876) and mFscore (0.887) higher than other models. In the face of unbalanced wetland categories, combining CrossEntropyLoss and FocalLoss in the loss function can improve several indicators of difficult cases to be segmented, enhancing the classification accuracy and generalization ability of the model. Finally, the category scale pie chart of Yancheng Binhai wetlands was plotted. In conclusion, this study achieves an effective segmentation of Yancheng coastal wetlands based on the semantic segmentation method of deep learning, providing technical support and reference value for subsequent research on wetland values.

Keywords: deep learning; semantic segmentation; SegFormer; coastal wetland; remote sensing images; machine learning

Citation: Lin, X.; Cheng, Y.; Chen, G.; Chen, W.; Chen, R.; Gao, D.; Zhang, Y.; Wu, Y. Semantic Segmentation of China's Coastal Wetlands Based on Sentinel-2 and Segformer. *Remote Sens.* **2023**, *15*, 3714. https://doi.org/10.3390/rs15153714

Academic Editors: Sawaid Abbas, Janet E. Nichol, Faisal M. Qamer and Jianchu Xu

Received: 30 May 2023
Revised: 18 July 2023
Accepted: 20 July 2023
Published: 25 July 2023

1. Introduction

The coastal wetland serves as an intermediary zone connecting marine and terrestrial ecosystems. It encompasses a vast expanse of coastal regions, intertidal areas and various aquatic environments, such as river systems, estuaries, salt marshes and sandy beaches. As a region of exquisite intricacy and ecological sensitivity, it undergoes profound influences from both the vast sea and the adjoining land. When it comes to the assessment of ecological service value, wetland ecosystems triumph over other ecosystems, such as oceans and forests. The second national wetland resources survey reveals that wetlands encompass a notable 5.58% of China's territory, with coastal wetlands alone constituting 10.85%. While

the coastal wetland occupies less than 1% of the entire country's land area, it plays a vital role in purifying polluted water, mitigating floods, minimizing the dangers of storm surges and hurricanes, and providing a favorable natural environment.

The assessment of ecosystem service value serves as a pivotal bridge connecting ecosystem research and management decision-making. Its primary objective is to comprehensively comprehend the current status and dynamic trends of ecosystem services. The creation of a wetland map plays a crucial role in establishing the indispensable spatial and geographical foundation for the quantitative evaluation and calculation of wetland ecological service value. It also facilitates a comprehensive understanding of the position and function of coastal wetlands within the broader ecosystem, encompassing factors, such as hydrological characteristics and species diversity. Furthermore, the generation of a wetland map provides fundamental data and information support for the delineation of wetland protection zones and the implementation of wetland protection and restoration measures. This invaluable resource enables decision-makers to holistically consider the ecological value of wetlands and averts the destruction and wastage of valuable wetland resources.

High-resolution remote sensing images encompass a diverse array of distinctive attributes, spanning spectral, size, structure, shape, geometric factors, layout and other relevant characteristics. The spectral characterization of these images results in heightened spectral variation among similar features, reduced inter-class variation and an amplification of homospectral and heterospectral phenomena. As a consequence, the classification of high-resolution remotely sensed imagery has become increasingly intricate. The advancements in remote sensing technology have made a substantial impact on wetland research, offering an effective means to assess, monitor and manage wetland ecosystems. This technology facilitates the acquisition of high-resolution, multi-temporal and multi-spectral remote sensing image data specifically tailored to wetlands. By combining digital image processing with machine learning techniques, the automatic extraction and analysis of wetland vegetation, hydrological patterns, soil characteristics and meteorological information can be accomplished. This integration significantly enhances the efficiency and precision of obtaining valuable wetland information.

Traditional wetland extraction methods primarily employ manual visual interpretation techniques. This approach involves professionals acquiring specific target feature information on remote-sensing images through direct judgment or with the assistance of auxiliary interpretation instruments. For example, grayscale co-occurrence matrix [1], normalized difference index and scale invariant feature transformation are used [2]. The task of visual interpretation can be both arduous and time-consuming for those who interpret images. In addition, the quality of interpretation is limited by the interpreters' experience and knowledge of the area while the time required for information acquisition can be lengthy. For these reasons, visual interpretation is not a suitable method for independent monitoring of wetlands on a larger spatial scale. To address this issue, some scholars have taken steps to improve upon this method. Aaron Judah et al. [3] proposed a method to improve wetland classification based on multi-source remote sensing data. Sam Jackson et al. [4] used hyperspectral and lidar data to analyze coastal wetland vegetation. M. Amani et al. [5–7] used Google satellites to classify the wetlands.

As computer technology continues to advance and the use of multiple sources of data becomes increasingly prevalent, machine learning algorithms based on probabilistic statistics are supplanting traditional supervised and unsupervised classification methods as a more effective means of wetland classification. U. Maulik et al. [8] used semi-supervised SVM to classify wetland remote sensing images. Kazi Rifat Ahmed et al. [9] proposed a new method of wetland classification using optical index and machine learning. B. Fu et al. [10] mapped wetland vegetation with object- and pixel-based random forest algorithms. J. Munizaga, Ali Gonzalez-Perez and Ricardo Martínez Prentice [11–13] classified wetlands using machine learning and high-resolution imagery. Zou and Li [14,15] used landsat8 satellite imagery to achieve wetland classification. Tian et al. [16] used random forest to

classify the vegetation of wetlands in arid areas of Xinjiang. L. F. Ruiz et al. [17–21] used Sentinel-1 images to classify wetland types.

Traditional classification methods based on spectral features, such as the maximum likelihood method, the minimum distance method and the K-mean clustering method, have suffered from reduced accuracy. Shallow learning algorithms, such as support vector machines and neural networks, are limited in their ability to effectively express complex functions and have restricted computational units. As the number of training samples and sample types increases, shallow models become ineffective in achieving desired classification results.

In recent times, deep learning algorithms have emerged as a promising approach for wetland information extraction, owing to their ability to abstract high-level features of images and efficiently reduce dimensionality. This development has been made possible by the advancements in image pattern recognition and artificial intelligence technology. B. Hosseiny et al. [22] proposed a spatiotemporally integrated deep learning model for wetland classification. Ali Jamali et al. [23–25] used Sentinel-1 and Sentinel-2 data to classify wetlands through Swin Transformer. Xingwei He et al. used Swin Transformer Embedding UNet for semantic segmentation of remote sensing images [26]. Z. Lv et al. used simple reading UNet for remote sensing image change detection [27]. Atharva Sharma et al. [28] applied a block-based convolutional neural network to extract roads and buildings from high-resolution remote sensing images simultaneously. Xin Li et al. [29] proposed a semantic segmentation network with dual attention depth fusion for large-scale satellite remote sensing images. Asif Raza et al. [30] used UNet++ to realize the change detection of high-resolution remote sensing images. Emmanuel Maggiori et al. [31] classified large-scale remote sensing images by the convolutional neural network. Transformer is a deep learning architecture commonly used in the field of natural language processing. It has also achieved some success in computer vision, which also includes the classification of some remote-sensing images [32–35].

The Yancheng Binhai Wetland, located in Jiangsu Province, stands as the first Binhai Wetland World Natural Heritage Site in China. In this piece of literature, we designate this area as our study site and employ Sentinel-2 to obtain high-resolution remote sensing data. We also utilize the SegFormer model, which is based on transformer architecture, to facilitate our wetland classification study. The contributions of our research are as follows: (1) We identify a set of feature variables that are suitable for the classification of coastal wetlands within China. (2) We summarize existing research and propose a comparative experimental method that can be applied to the classification of coastal wetlands. (3) We accomplish the semantic segmentation of coastal wetlands situated in Yancheng. (4) We put forward a loss function that combines CE_Loss and FocalLoss to address the commonly encountered issue of class imbalance in wetlands. This approach enhances the classification accuracy and generalization ability of our model. (5) We create a pie chart, which depicts the class proportions of coastal wetlands in Yancheng, for future wetland value research calculation, facilitating it as a point of reference.

2. Study Area and Data Preprocessing

2.1. Study Area

Yancheng Wetland is situated in the eastern coastal region of Jiangsu Province, China, as depicted in Figure 1. This wetland is the largest coastal wetland on the west coast of the Pacific Ocean and the edge of the Asian continent, with a sprawling area of approximately 4533 km^2. It constitutes 7/10 of the total mudflat area in Jiangsu Province and 1/7 of the nation's total mudflat area. This wetland has been included in the World Key Wetlands List and is renowned as the "Wetland Capital of the East". Yancheng coastal wetlands are chiefly located in the middle of the coast of northern Jiangsu Province, spanning the coastal regions of Ringshui, Binhai, Sheyang, Dafeng and Dongtai County. These wetlands can be found between 32°32–34°25 north latitude and 119°55–121°50 east longitude. The coastline of Yancheng extends 580 km, accounting for 2/3 of the entire coastline of Jiangsu Province,

and the coastal mudflats cover an area of 4500 km^2. Based on preliminary calculations, the wetland area of Yancheng covers about 20% of the entire area of Yancheng, and it features two national key nature reserves. Yancheng wetlands are classified under the subtropical monsoon climate zone, experiencing hot and rainy summers and cold and dry winters. The average annual temperature ranges from 15–17°, and the annual precipitation amounts to roughly 1000 mm. Owing to tidal action, the mudflats within the area are continuously being silted up, resulting in the formation of a coastal wetland plain that is home to a diverse range of wetland types. The primary types of wetland vegetation cover include reeds, sod and rice grass.

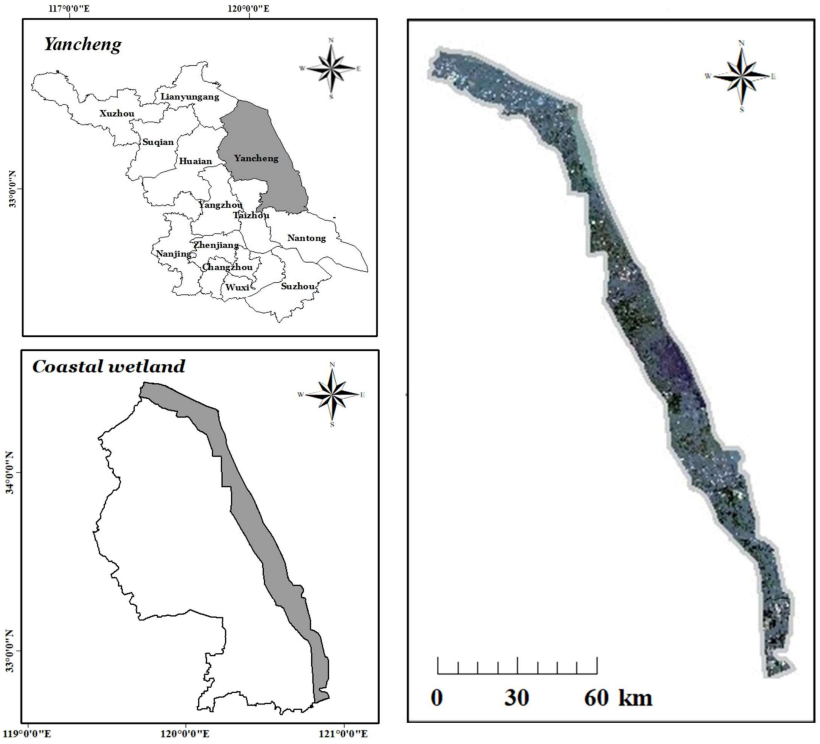

Figure 1. Yancheng Coastal Wetland, Jiangsu Province, China.

2.2. Data Collection

This study utilizes remote sensing images from the Sentinel-2 satellite of the Copernicus Data Centre of ESA, captured in August and September, as the data source for the Binhai wetland in Yancheng, Jiangsu Province. Sentinel-2 provides high spatial resolution (10 m) and multiple bands (13 bands) of data, encompassing vegetation, soil and water cover, inland waterways and coastal areas. For this research, remote sensing images captured during the summer season were selected. This choice was made because the hydrological characteristics of coastal wetlands are more distinct during this period, and the color variations of different wetland vegetation types in remote sensing images are more evident. These factors aid in manual visual labeling and other related tasks. Figure 2 presents an overview of the complete data processing procedure. We used software, such as SNAP and ENVI, to preprocess the source Sentinel data. This involved a sequential process of atmospheric correction, resampling, band synthesis and vector cropping to obtain the initial processed remote sensing images.

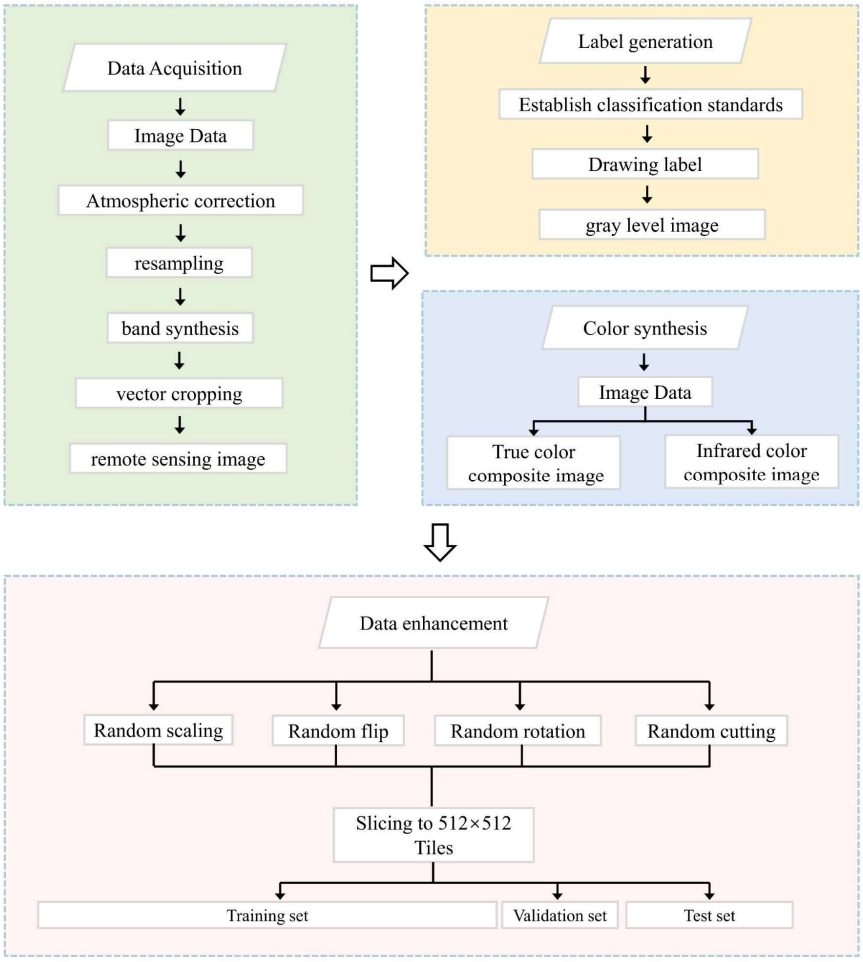

Figure 2. The flow chart of data processing.

As previously mentioned, we utilized snap and envi software to preprocess the source Sentinel data for this study. Atmospheric correction, resampling, band synthesis and vector cropping were sequentially employed to obtain the preliminary processed remote sensing images.

2.3. Wetland Classification

The concept of wetlands is comprehensive, encompassing natural or artificial, permanent or temporary marshes, peatlands or water areas, static or flowing, and can contain freshwater, brackish or saltwater bodies, including waters up to 6 m deep at low tide. Wetlands can be classified into various types, which can be broadly categorized into natural and artificial wetlands. For this study, we developed a wetland classification system by integrating previous research and utilizing existing satellite images. As presented in Table 1, the wetland classification system comprises eight types of wetlands, including Construction land, *S. alterniflora*, *S. salsa*, *P. australis*, Farmland, River system, Aquaculture and Tidal flat.

2.4. Production of Data Sets

For this study, we utilized Sentinel-2 satellite images with a resolution of 10 m. The Yancheng Binhai wetland is distributed along the coastline as a whole and has a strip shape with a size of 14,516 × 20,956. We initially cropped the large image into 10 images with a size of 2922 × 2247, based on the strip distribution characteristics of the wetland. Additionally, we manually labeled the images through visual inspection, in accordance with the wetland classification system mentioned previously. The specific categories corresponding to colors are presented in Table 2.

Table 1. Wetland classification criteria.

First Classification	Secondary Classification	Remote Sensing Image	Geometric Feature
Natural wetland	*S. alterniflora*		Green, single texture, redundancy and continuity
	S. salsa		Pale yellow, flat, longitudinal stripes
	P. australis		Brown, green, monotonous texture
	River system		Light blue, striped, slender
	Tidal falt		Yellow, striped and distributed along the coast.
Artificial wetland	Farmland		Green, striped, distributed along the river.
	Aquaculture		Light color, regular grid, distributed along lakes.
Non-wetland	Construction land		Light gray, light yellow, textured and regular in shape.

Table 2. Wetland Category Colors.

Class Name	Color	Color Channel (R, G, B)	Color Channel (Gray)
Construction land	red	200, 0, 0	1
S. alterniflora	gold yellow	250, 200, 0	2
S. salsa	olive	200, 200, 0	3
P. australis	lime green	150, 250, 0	4
Farmland	green	0, 200, 0	5
River system	blue	0, 0, 200	6
Aquaculture	light blue	0, 150, 200	7
Tidal falt	deep pink	200, 150, 150	8
Background	white	255, 255, 255	0

2.5. Data Enhancement and Data Filtering

Our dataset comprises images in four bands, namely red (0.63–0.69 μm), green (0.52–0.59 μm), blue (0.45–0.52 μm) and near-infrared (0.77–0.89 μm). We generated true color synthetic images and NIR synthetic images by utilizing different band combinations. To enhance the training speed of our dataset, we converted RGB three-channel labels into single-channel labels and employed grayscale maps as the labels for training. Figure 3 displays our trained remote-sensing images and the corresponding labels.

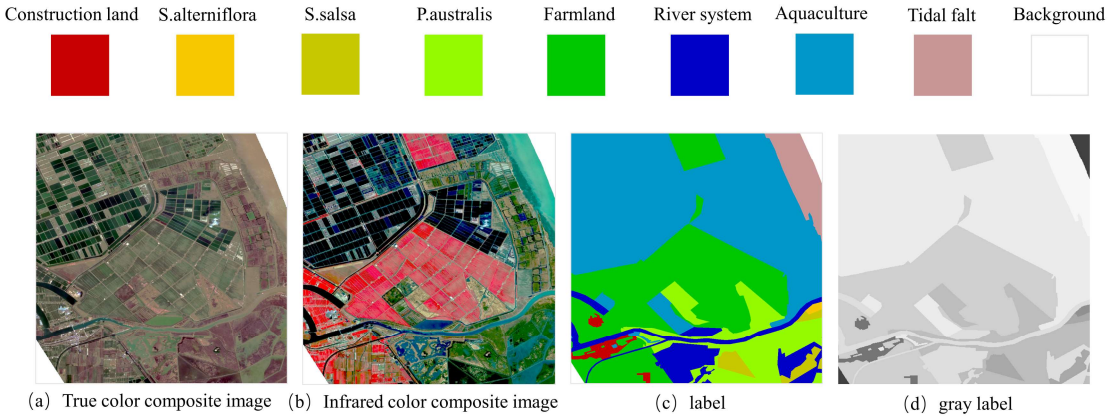

Figure 3. Remote sensing images and labels in the training set.

In the literature [36–39], the authors have shown that image cropping can improve the local effect of prediction. To augment the dataset, we initially apply random zooming and rotation/flip operations to the images in the dataset, followed by random cropping based on sizes of 128 × 128, 256 × 256, and 512 × 512. Additionally, we exclude images with excessive background and retain those containing some boundary information, ultimately accumulating 15,443 samples, which we subsequently adjust to a uniform size of 512 × 512. We partition the entire dataset into a training set, validation set, and test set in a 7:1:2 ratio, resulting in the final dataset.

3. Method

3.1. Model

Enze Xie et al. introduced the SegFormer model in 2021, which is an image segmentation model based on Transformers [40]. In contrast to conventional convolutional neural networks, SegFormer exhibits superior scalability and adaptability by employing multiple layers of Transformer encoders for feature learning. By conducting multi-layer feature extraction and employing context awareness of images, SegFormer can proficiently comprehend the semantic information in images, and thus, generate precise segmentation outcomes.

SegFormer comprises two primary modules, as illustrated in Figure 4. The first module is a hierarchical Transformer encoder that generates both high-resolution coarse features and low-resolution fine features. The second module is a lightweight ALL-MLP decoder that fuses these multi-level features to generate the final semantic segmentation masks.

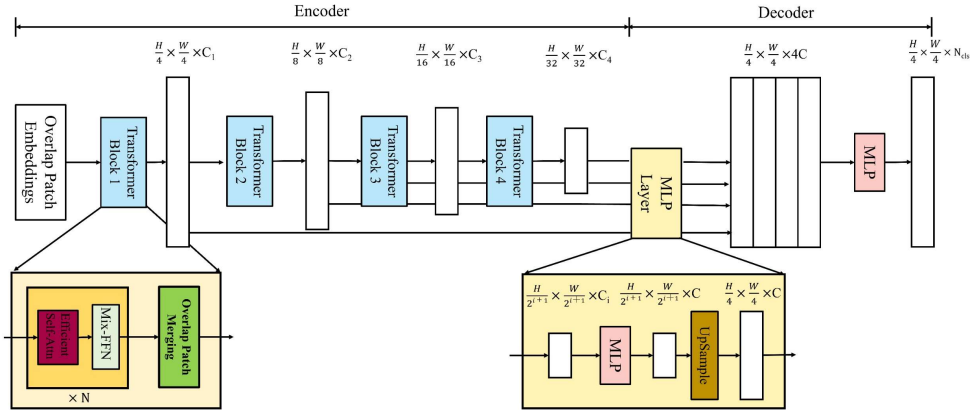

Figure 4. Model structure.

3.1.1. Encoder

The primary component of the encoder is a multi-layer transformer block. The Transformer module within the encoder, known as MiT, utilizes the overlap patch embeddings (OPE) structure for feature extraction and downsampling of the input image. The resulting features are then fed into the Multi-headed Self-attention layer and the Mix-FFN layer. Let us elaborate on these aforementioned structures.

3.1.2. Overlapped Patch Embeddings

In the SegFormer architecture, the overlap patch embedding module is utilized to extract image features. This module partitions the input image into multiple overlapping patches and generates an embedding vector for each of them. The overlap region between adjacent patches provides a broader perceptual field for capturing more contextual information while preserving the local information of each patch. This approach reduces information loss and mitigates the issue of boundary artifacts, thereby enhancing the robustness and generalization capability of the model.

3.1.3. Overlapped Patch Merging

In SegFormer, the Overlapped Patch Merging (OPM) technique is utilized to merge overlapping patches into a complete image. This module is inspired by ViT and is designed to combine non-overlapping images or feature blocks in ViT. Chunking may produce boundary effects, resulting in segmentation outcomes with discontinuous regions due to the absence of shared information between cropped blocks. Therefore, OPM leverages an overlapping patch merging process to eliminate these discontinuous edges by fusing adjacent blocks with similar or identical category labels to ensure spatial semantic coherence. The module can be implemented using a convolutional layer with Kernel = 7, Stride = 4 and Padding = 3.

3.1.4. Efficient Self-Attention

Efficient Self-Attention is a lightweight self-attention mechanism utilized for pixel-level feature extraction. In comparison to the conventional Self-Attention approach, it minimizes the computational load and number of parameters, thereby improving the model's efficiency during runtime and mitigating the issue of overfitting.

In the original multi-head self-attention process, each of the heads Q, K, V has the same dimensions $N \times C$, where $N = H \times W$ is the length of the sequence, the self-attention is estimated as:

$$Attention(Q, K, V) = Soft\ max(\frac{QK^T}{\sqrt{d_{head}}})V \tag{1}$$

where Q is the query vector, K is the key vector and V is the value vector. d_{head} is the latitude size of each vector, and Softmax is a normalization function.

The traditional Self-Attention needs to calculate the similarity matrix between any two positions during the computation, and the computational complexity of this process is $O(N^2)$, which generates a large computation and memory occupation and cannot satisfy the input of large-resolution images. To reduce the complexity, ESA adopts an idea similar to local attention and introduces the approximate ratio R to approximate the length of the input sequence.

$$\hat{K} = Reshape(\frac{N}{R}, C \cdot R)(K) \tag{2}$$

$$K = Linear(C \cdot R, C)(\hat{K}) \tag{3}$$

where Reshape function converts K into a sequence of $\frac{N}{R} * (C - R)$ shapes, Linear transforms $\hat{K}$ linearly and finally gets K of dimension $\frac{N}{R} * C$. The complexity of the ESA mechanism is reduced from $O(N^2)$ to $O\left(\frac{N^2}{R}\right)$. In the 4 stages of the model, R is set as (64,16,4,1).

3.1.5. Mix-FFN

Conventional Vision Transformer (ViT) uses Position Encoding (PE) to introduce position information. However, the resolution of PE is fixed. Therefore, when the test resolution and the training resolution are different, the position encoding needs to be interpolated, but it often leads to a decrease in accuracy. The Mix-FFN, the structure used in SegFormer, is shown in Figure 5. It takes into account the effect of zero padding on the leaked position information and uses 3×3 Conv directly in the feedforward network (FFN), whose equation is shown below.

$$X_{out} = MLP(GELU(Conv_{3\times3}(MLP(X_{in})))) + X_{in} \tag{4}$$

where X_{in} is the feature from the self-attention module. Mix-FFN mixes a 3×3 convolution and an MLP into each FFN.

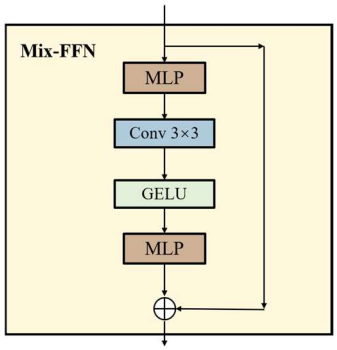

Figure 5. Mix-FFN.

3.2. Lightweight All-MLP Decoder

The Lightweight All-MLP Decoder is a compact decoder that processes the feature maps generated by the encoder output to obtain the final semantic segmentation outcome. This decoder has a reduced number of parameters, and the multi-level Transformer Encoder can obtain a broader perceptual field than the conventional CNN Encoder.

The proposed All-MLP decoder consists of four main steps. First, multi-level features F_i from the MiT encoder go through an MLP layer to unify the channel dimension. Then, in a second step, features are up-sampled to 1/4th and concatenated together. Third, an MLP layer is adopted to fuse the concatenated features F. Finally, another MLP layer takes the fused feature to predict the segmentation mask M with a $\frac{H}{4} \times \frac{W}{4} \times N_{cls}$ resolution, where N_{cls} is the number of categories. The formula is shown below.

$$\hat{F}_i = Linear(C_i, C)(F_i), \forall i \tag{5}$$

$$\hat{F}_i = Upsample(\frac{W}{4} \times \frac{W}{4})(\hat{F}_i), \forall i \tag{6}$$

$$F = Linear(4C, C)\left(Concat\left(\hat{F}_i\right)\right), \forall i \tag{7}$$

$$M = Linear(C, N_{cls})(F) \tag{8}$$

where M refers to the predicted mask and $Linear(C_{in}, C_{out})(\cdot)$ refers to a linear layer with C_{in} and C_{out} as input and output vector dimensions, respectively.

3.3. Model Training

In this experience, the gradient descent algorithm with momentum (Adam with weight decay, adamw) is chosen as the optimization algorithm for model training, and the model parameters are updated by minimizing the difference between the predicted

and real features until the loss function reaches the minimum value. The momentum gradient descent method borrows the concept of momentum in physics when updating the parameters so that the update of parameters not only depends on the currently calculated gradient information but also takes into account the gradient information of the previous update steps. By using the exponentially weighted momentum m_t to replace the original gradient g_t for updating the model parameters, the model can avoid the oscillation problem during the training process and speed up the training of the model. m_t and the parameter update are calculated as shown in Equations.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{9}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) {g_t}^2 \tag{10}$$

where g_t is the gradient of the current step. m_t is the first-order moment estimate of the gradient. v_t is the second-order moment estimate of the gradient. β_1 and β_2 are the corresponding exponential decay rates, respectively.

$$\hat{m}_t = \frac{m_t}{1 - (\beta_1)^t} \tag{11}$$

$$\hat{v}_t = \frac{v_t}{1 - (\beta_2)^t} \tag{12}$$

$$\alpha_t = \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} \tag{13}$$

where η is the initial learning rate. ε is a very small constant to avoid the error of dividing by zero.

$$w_{t+1} = w_t - \alpha_t (\hat{m}_t + \lambda w_t) \tag{14}$$

where λ is the weight decay factor.

The deep learning server used in this study had a specific hardware and software configuration, which is detailed in Table 3. To ensure consistency, other experimental parameters were standardized, as shown in Table 4, to evaluate the models' performance on the dataset. The momentum descent algorithm with gradient was utilized, and the adamw optimizer, which incorporates the weight decay mechanism, was found to be faster and more accurate than other gradient descent algorithms with minimal human intervention. The CrossEntropyLoss function was used as the loss function for the experiments. Additionally, a preheating mechanism was implemented to compare the strengths and limitations of each model by setting the maximum Epoch to 70.

Table 3. The environment of the experiment.

GPU	**Video Memory**	**Operating System**	**Language**
NVIDIA GeForce RTX 3090	24 G	Linux	Python 3.9
CPU	**Random Access Memory**	**Deep Learning Framework**	**Cuda**
Random Access Memory Intel (R) Xeon (R) Gold 6330	60	Pytorch 1.7.0	Cuda 11.0

Table 4. Network parameters of the considered architectures.

Optimizer	Momentum	Learning Rate (LR)	LR Scheduler	Weight Decay	Loss Function	Warmup
AdamW	0.9	0.0001	Polynomial decay	0.01	CrossEntropyLoss	linear
Warmup Iters	**Warmup Ratio**	**Iters**	**Batch Size**	**Max Epoch**	**Validation Frequency**	
1500	1×10^{-6}	29,190	8	70	Each epoch	

3.4. Evaluation Metrics

To assess the semantic segmentation performance, we employed several widely-used metrics, including Mean Accuracy (*mAcc*), Overall Accuracy (*Acc*) and Mean Intersection over Union (*mIoU*).

$$mAcc = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FP_i} \tag{15}$$

$$aAcc = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n}(TP_i + FN_i)} \tag{16}$$

where n is the total number of categories, TP_i is the number of samples that the model predicts as class i and are truly labeled as class i, FN_i is the number of samples that are class i but are wrongly predicted as other classes and FP_i is the number of samples that the model predicts as class i for all other classes.

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ij} - p_{ii}} \tag{17}$$

where p_{ii} represents the number of pixels predicted by category i to category i. $k+1$ is the total number of categories, and p_{ij} is the number of pixels predicted by category i to category j.

The precision, recall, F_1-score and intersection over union (*IoU*) metrics are used to further evaluate the performance of various models on the test data set, which is calculated as follows.

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F_1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

where *TP* denotes the true positives, *FP* denotes the false positives and *FN* denotes false negatives. Based on these three metrics, we applied independent test set data to evaluate the performances with different models and different loss functions.

4. Experiments and Results

Results with Different Models

The same set of parameters was utilized on the test set to evaluate the performance of the models, and the results are depicted in Figure 6. SegFormer exhibited superior accuracy and robustness in multi-categorical semantic segmentation, as evidenced by its curve (yellow) outperforming other models across all metrics. At specific points, the

SegFormer model achieved a mean Intersection over Union (mIoU) of 0.81, mean Accuracy (mAcc) of 0.87, and average Accuracy (aAcc) of 0.94. In contrast, the Unet model (orange) demonstrated poorer performance across various metrics.

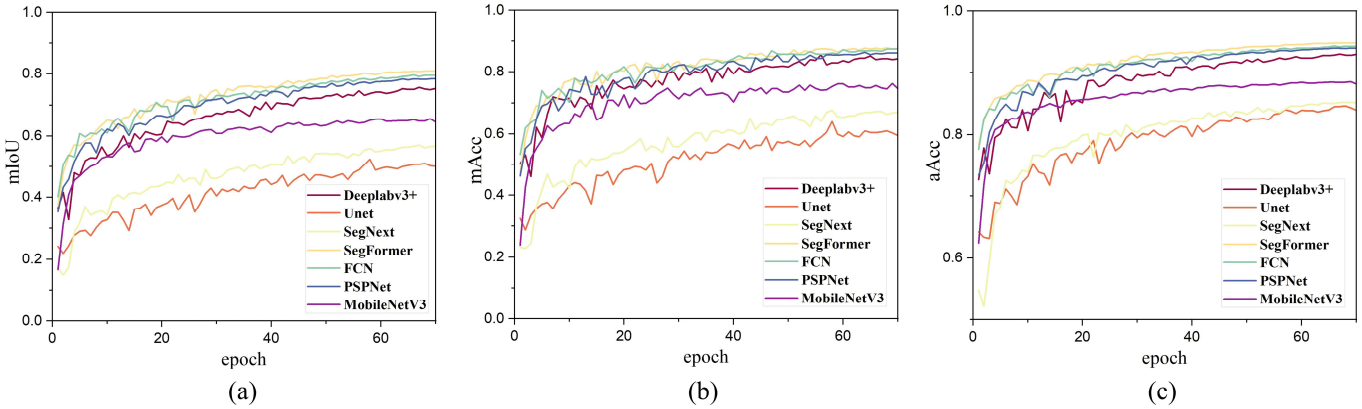

Figure 6. Comparison of the metrics of various models in model training (one validation per epoch). (**a**) The graph of mIoU. (**b**) The graph of mAcc. (**c**) The graph of aAcc.

Figure 7 shows the refined metrics of SegFormer for each class. As illustrated in Figure 7a, the model converges at approximately epoch 80. Figure 7b showcases some variations in the prediction outcomes for different categories, with construction land, s.alterniflora and farmland exhibiting better prediction results, with an IoU close to 90%. The river system's prediction effect is not satisfactory, with an IoU of about 40%. Figure 7c indicates that the overall prediction performance is good, with minor fluctuations in the Acc curve but in an upward trend. The accuracy of most categories can reach 90%, with the river system being the least accurate, with an accuracy of 50%.

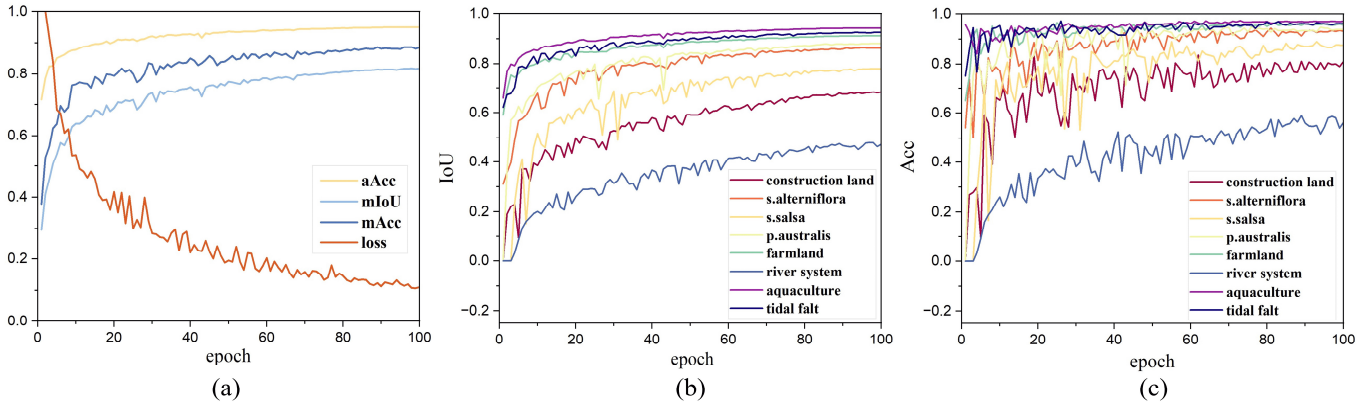

Figure 7. The evaluation results of segformer. (**a**) The iterative graph of aAcc, mIoU, mAcc and loss. (**b**) IoU curve for different classes. (**c**) Acc curve for different classes.

To further assess the model's classification effectiveness, Precision, Recall and F_1-score were selected, and the final test results on the test dataset are presented in Table 5. The table illustrates different prediction effects for each class, with significant differences in the indices for each model. SegFormer demonstrates the highest mPrecision (0.901) and the lowest (0.754), the highest mRecall (0.876) and the lowest (0.642), and the highest mF_1-score (0.887) and the lowest (0.632). These results indicate that SegFormer is the optimal model for remote sensing wetland classification, accurately identifying various categories of pixels and exhibiting better recognition of the shape and contour of the target object with high accuracy and consistency. Most categories achieved good results across all metrics in each model, but a few classes exhibited poor prediction results. Among the nine categories,

the river system consistently had low indices, with the highest Recall being only 0.549. Therefore, the river system proves to be a challenging category in wetland classification.

Table 5. The comparison between various models in terms of Precision, Recall and F-1 score (CL = Construction land, *S. alterniflora* = S.a, *S. salsa* = S.s, *P. australis* = P.a, Farmland = F, River system = RS, Aquaculture = A, Tidal falt = T, Background = B, mRecall = MR, mPrecision = MP and mF-score = MF).

Model	CL	S.a	S.s	P.a	F	RS	A	T	B	MP	MR	MF
Deeplabv3+										0.859	0.843	0.850
Precision	0.716	0.881	0.848	0.913	0.927	0.626	0.962	0.937	0.919			
Recall	0.727	0.897	0.762	0.910	0.946	0.507	0.954	0.945	0.935			
F-1 score	0.722	0.889	0.803	0.911	0.937	0.560	0.958	0.941	0.927			
Unet										0.837	0.798	0.826
Precision	0.698	0.875	0.805	0.870	0.866	0.744	0.875	0.926	0.928			
Recall	0.746	0.840	0.745	0.817	0.869	0.524	0.942	0.824	0.933			
F-1 score	0.721	0.857	0.774	0.843	0.867	0.618	0.908	0.872	0.930			
SegNext										0.754	0.672	0.695
Precision	0.574	0.764	0.552	0.799	0.834	0.616	0.890	0.885	0.872			
Recall	0.342	0.730	0.555	0.679	0.903	0.164	0.904	0.873	0.901			
F-1 score	0.429	0.746	0.554	0.734	0.867	0.259	0.897	0.879	0.886			
SegFormer										**0.901**	**0.876**	**0.887**
Precision	0.791	0.924	0.917	0.923	0.948	0.745	0.968	0.957	0.933			
Recall	0.781	0.929	0.841	0.947	0.956	0.541	0.970	0.960	0.965			
F-1 score	0.786	0.926	0.877	0.935	0.952	0.627	0.969	0.959	0.948			
FCN										0.892	0.870	0.879
Precision	0.798	0.910	0.904	0.921	0.937	0.729	0.971	0.946	0.908			
Recall	0.751	0.926	0.826	0.940	0.957	0.549	0.961	0.958	0.960			
F-1 score	0.774	0.918	0.863	0.930	0.947	0.626	0.966	0.952	0.933			
PSPNet										0.879	0.864	0.869
Precision	0.721	0.893	0.898	0.922	0.938	0.702	0.965	0.946	0.923			
Recall	0.769	0.927	0.810	0.927	0.947	0.531	0.961	0.952	0.951			
F-1 score	0.744	0.910	0.852	0.924	0.943	0.604	0.963	0.949	0.937			
MobileNetV3										0.792	0.761	0.771
Precision	0.601	0.812	0.696	0.848	0.867	0.563	0.924	0.902	0.915			
Recall	0.566	0.800	0.745	0.845	0.912	0.277	0.917	0.901	0.882			
F-1 score	0.583	0.806	0.720	0.846	0.889	0.372	0.921	0.901	0.898			

In order to compare the accuracy and stability of different models more intuitively, we selected test set images for prediction. Figure 8 showcases a remote sensing image of the Yancheng Wetland Rare Birds National Nature Reserve in Jiangsu Province, which represents a typical coastal wetland. This area is characterized by a distinct estuarine delta landscape and consists of various types of wetland environments, including freshwater marshes, lakes, beaches and coastal wetlands of the Yangtze River Delta. It is home to a diverse range of species and holds high ecological conservation value.

The predicted results are presented in Figure 8. Combining the images and labels, the distribution of *S. alterniflora*, *S. salsa*, *P. australis* and Tidal falt in this reserve is relatively concentrated, accounting for most of the reserve. The identification of these four categories was generally consistent among the models, and all models except Unet were able to extract clear boundary information. Deeplabv3+, SegFormer, FCN, Unet and PSPNet were able to identify the river system at the edge of the protected area while the other models, SegNext and MobileNetV3, were able to extract the river system relatively accurately. However, the MobileNetV3 extraction results are finer than the actual ones. It is observed that the SegNext model exhibits inaccuracies in differentiating between paddy fields, ponds and rivers in arable land, leading to misclassifications of the River system and Aquaculture categories. Notably, the part marked by the larger red box in the labeled image is the scattered River system. In the prediction of this part, most models' segmentation results produce too many isolated points, with relatively smooth and lacking contour features. In contrast, SegFormer significantly outperforms other methods in terms of result accuracy and edge clarity. Additionally, in the other red box, most models ignore the river due to its small and slender area, with only SegFormer extracting a portion of the water body.

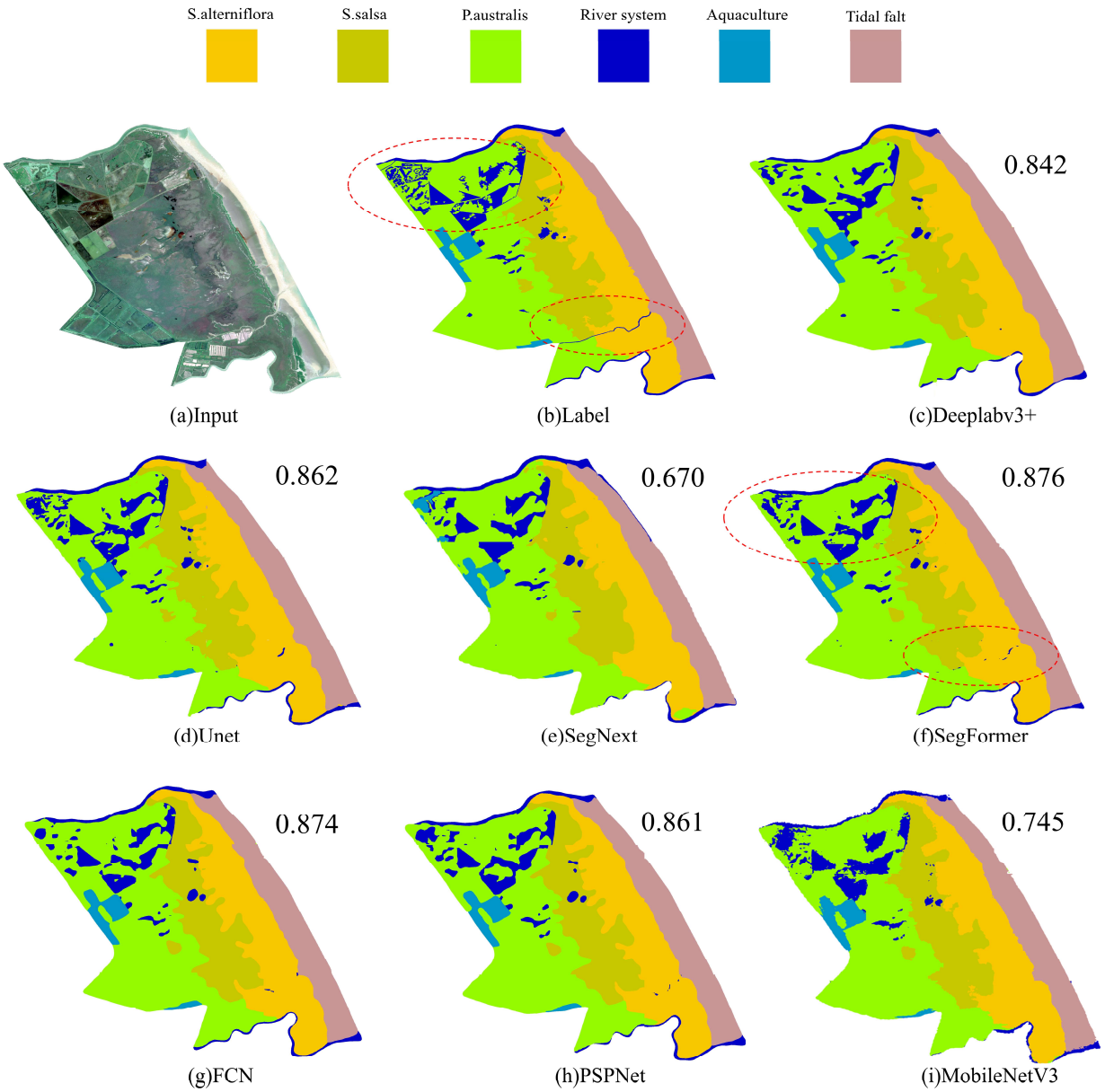

Figure 8. A comparison image about the predicted images of different deep learning models.

Figure 9 presents a comparison between the prediction results of deep learning methods and traditional methods, including supervised classification based on the statistical maximum likelihood method, machine learning-based neural networks and supervised classification based on support vector machines (SVM). While the traditional methods mentioned above may outperform deep learning in extracting boundary information, they suffer from misclassification issues. For example, the maximum likelihood method misclassifies *P. australis* as *S. salsa* while the neural network exhibits the opposite misclassification. SVM, on the other hand, struggles to accurately differentiate the river system from Aquaculture. These findings highlight the limitations of traditional methods when it comes to precise classification in complex wetland environments.

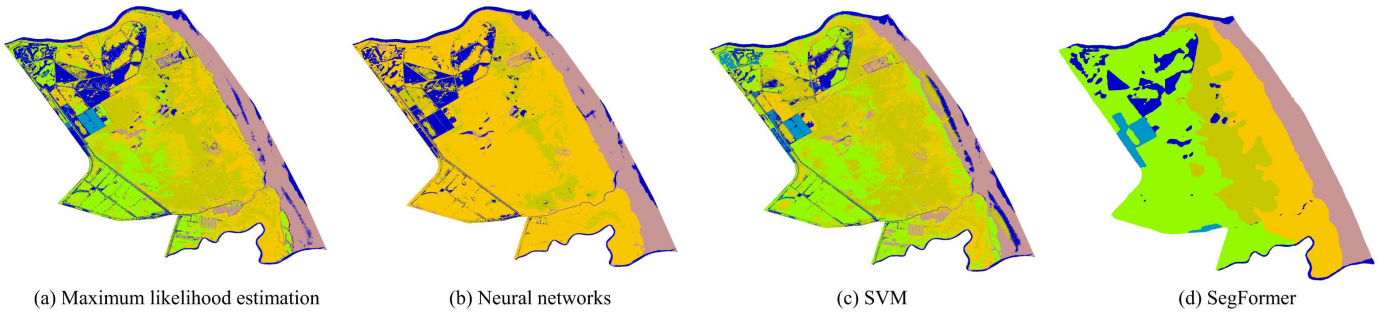

Figure 9. Prediction graphs compared to traditional methods.

In order to further investigate the classification capability of the deep learning model mentioned above for small-scale features, we conducted additional comparisons using various sets of images, as illustrated in Figure 10. These images contained a diverse range of feature characteristics. From the prediction results, it can be inferred that the River system remains a challenging aspect of the prediction process. The SegFormer-based prediction method outperforms the others, as it can effectively extract most of the features in the water body. PSPNet and FCN methods are also effective but not as much as the SegFormer approach. In summary, the SegFormer model, based on deep learning, exhibits an impressive ability to perform semantic segmentation of wetland remote sensing images and is capable of learning and extracting the features inherent in wetlands. It can effectively segment boundaries in complex ground object environments and extract small-scale ground objects that are affected. As one of the deep learning methods, with the emergence of higher resolution images and more powerful computing devices, it is expected to exhibit even faster segmentation efficiency and higher accuracy than traditional methods.

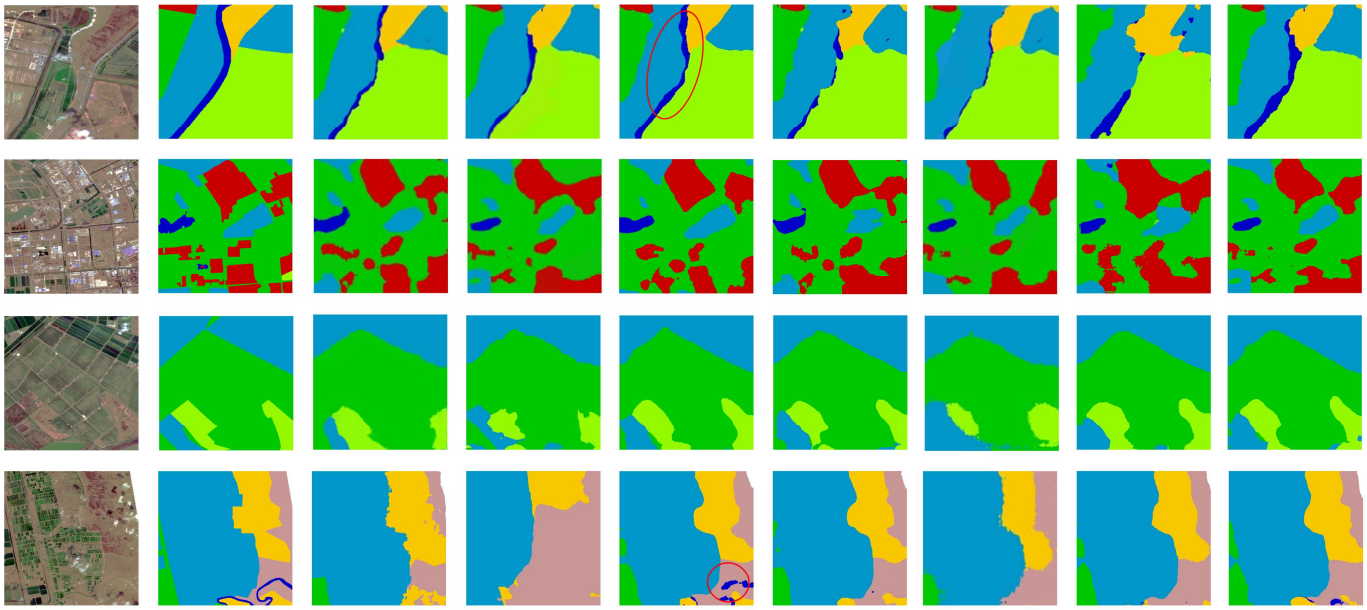

Figure 10. Predictive graphs of different models.

5. Discussion

5.1. Results with Different Loss Functions

As illustrated in Figure 11, there exists a specific class imbalance in the dataset. When considering the relationship between the proportion of features in the training samples and the accuracy of the model's extraction results, it was observed that the extraction accuracy was higher for the feature classes that were more adequately rep-

resented in the dataset, such as Tidal falt, Aquaculture, Farmland, *P. australis* and *S. alterniflora*, for each model. Conversely, the extraction accuracy was relatively poor for categories that accounted for a smaller proportion of the river system, such as *S.salsa* and Construction land dataset. Due to the limited information contained in the samples for these categories, it becomes difficult for the network to learn their distribution pattern, and it is challenging to effectively identify them.

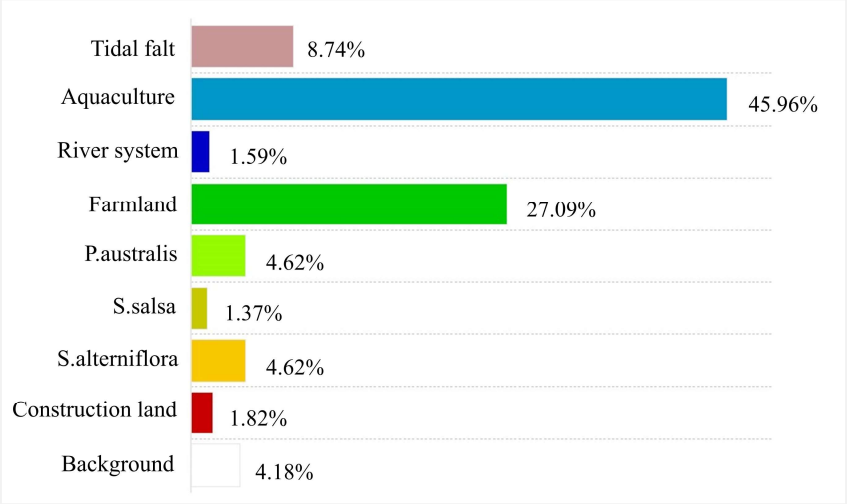

Figure 11. The proportion of the Number of Pixels (%).

In order to improve the extraction accuracy of the model, we explored multiple loss functions. Specifically, we chose the best model, SegFormer, and modified only the loss function while keeping the other parameters unchanged. For this study, we considered CrossEntropyLoss (CE_Loss), LovaszLoss, Focal Loss and CrossEntropyLoss + FocalLoss as loss functions, and the results are presented in Table 6. The findings in Table 6 indicate that, without any changes to the deep learning algorithm framework and optimization, LovaszLoss performs the worst, with its various metrics being lower than the other loss functions. FocalLoss is slightly inferior to CE_Loss, but it improves the performance of ConStruction land classification in terms of IoU, Acc and F-score metrics. The combination of CE_Loss and FocalLoss enhances the classification performance of the model under this problem. Specifically, the IoU and F-score for River system classification improved by 0.28% and 0.27%, respectively, compared to CE_Loss. Additionally, Acc for *S. salsa* classification improved by 0.56% while Acc and F-score for Construction land classification improved by 1.41% and 1.04%, respectively. Overall indexes, such as mIoU, mAcc, aAcc and mF-score, also improved by about 0.1% compared to CE_Loss. It can be concluded that, under the condition of unbalanced wetland classification, the combined loss function of CE_Loss and FocalLoss can enhance the classification accuracy of difficult classes, alleviate overfitting and misclassification of the model caused by class imbalance and improve the performance and generalization ability of the model.

5.2. Coastal Wetland Detection

Changes in precipitation and temperature can indeed have an impact on the landscape pattern of coastal wetlands. The gradual increase in temperature can lead to increased evaporation of surface water, reduced precipitation and limited availability of water resources. This puts additional strain on coastal wetlands and disrupts their ecological balance to some extent. Additionally, the growing population in Yancheng has contributed to the degradation of coastal wetlands. The demand for ecological resources provided by wetlands has risen, putting tremendous pressure on these fragile ecosystems. Excessive development and utilization have further worsened the situation, resulting in a significant decrease in the coastal wetland area. Considering the rapidly changing wetland environment, it is crucial

to implement efficient and timely detection methods. These methods will help monitor and assess the status of coastal wetlands effectively. By detecting changes in land cover, vegetation and other relevant parameters, we can gain valuable insights into the health and conservation needs of these ecosystems. Timely detection and monitoring enable us to take prompt action and implement appropriate management strategies to mitigate the negative impacts and ensure the long-term sustainability of coastal wetlands.

Table 6. Comparison of different loss functions.

Loss Function	IoU	Acc	F-Score	mIoU	mAcc	aAcc	MR	MP	MF
CE_Loss				81.05	87.64	94.84	87.64	90.06	88.65
River system	45.62	54.07	62.65						
S. salsa	78.16	84.08	87.74						
Construction land	64.71	78.07	78.57						
LovaszLoss				77.72	87.62	92.84	87.62	85.31	86.31
River system	39.08	55.99	56.2						
S. salsa	76.1	86.64	86.43						
Construction land	58.23	85.24	73.6						
FocalLoss				80.48	87.46	94.78	87.46	89.34	88.24
River system	43.91	53.44	61.02						
S. salsa	75.63	82.97	86.12						
Construction land	65.48	79.27	79.14						
CE_Loss + FocalLoss				81.15	87.73	94.91	87.73	89.99	88.73
River system	45.9	56.02	62.92						
S. salsa	77.25	84.64	87.16						
Construction land	66.12	77.16	79.61						

Traditional wetland inspection mainly relies on manual surveys and visual inspections of remote sensing images [41–43], which consume significant human and material resources. While various digital image processing methods have been proposed to improve the classification accuracy of detailed information in remote sensing images, they are often limited in their ability to meet the requirements of complex image classification [44–46]. Although machine learning methods have been shown to enhance classification accuracy [47–50], they require image preprocessing and feature enhancement to achieve better classification results [51]. Moreover, the degree of automation is not high, and significant improvements in the correct rate are already difficult to achieve. In this paper, we propose more specific methods and ideas for wetland classification, building upon existing deep learning methods and demonstrating their feasibility and research value.

In this study, we utilized high-resolution images provided by Sentinel-2 to count the percentage distribution of feature classes in the study area, enabling subsequent calculation of ecological values. The specific proportional information is illustrated in Figure 12.

Figure 12 reveals that Aquaculture, Farmland and Tidal flat constitute the predominant components of Yancheng's coastal wetland, accounting for 42.2%, 29.7% and 14.7% respectively. Notably, *S. salsa* and *P. australis*, which represent unique vegetation within the wetland, encompass just 1% and 3.6%, respectively, totaling an area of 44.79 km^2 and 100.84 km^2. The utilization of deep learning techniques enables the swift detection of changes in Yancheng's coastal wetland, as well as variations in the proportion of different species. The value of wetland ecosystem services serves as a monetized representation illustrating the collective contribution of wetland ecosystems to human well-being. It encompasses an array of services, such as product supply, ecological regulation, and spiritual and cultural benefits. When evaluating the wetland ecosystem service value of Yancheng, it is crucial to select appropriate evaluation indices and systems aligned with the region's available resources and market conditions. The wetland mapping provides valuable spatial and geographic foundations that facilitate effective assessment of the wetland ecosystem service value.

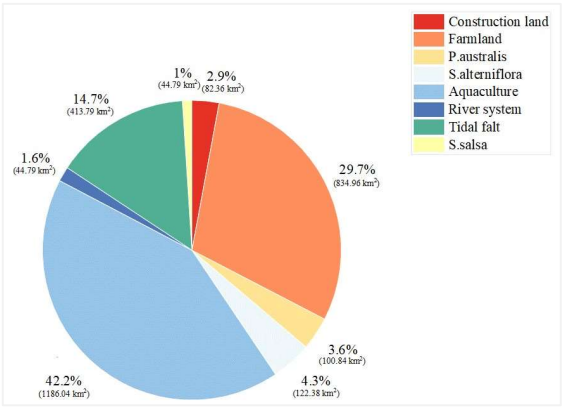

Figure 12. Pie chart of feature class scales for the study area.

Regarding the significant components of Yancheng's coastal wetland, namely Aquaculture, farmland and mudflat, their construction should prioritize enhancing the ecosystem service value concerning product supply, water flow regulation and water purification. As for the special vegetation occupying a comparatively smaller proportion, a comprehensive assessment of their ecosystem service value should concentrate on aspects related to species conservation, tourism and recreation, as well as scientific and educational services.

6. Conclusions

In this study, we utilized high-resolution images provided by Sentinel-2 and constructed an information extraction model of the Yancheng Binhai wetland based on SegFormer. Based on the relevant literature and remote sensing images, we classified the wetlands in the area into nine species, namely, Construction land, *S. alterniflora*, *S. salsa*, *P. australis*, Farmland, River system, Aquaculture and Tidal falt. Remote sensing images were preprocessed using atmospheric correction, resampling, band synthesis and vector cropping. The labeled images were drawn, and data enhancement methods, such as random deflation, random flip and random crop, were applied. Finally, the training images were standardized to 512 × 512 size and divided into a training set, validation set and test set in a ratio of 7:1:2. Evaluation metrics, such as mIoU, mAcc, aAcc, mRecall, mPrecision and mF-score were used. While ensuring the consistency of other parameters, we conducted comparison experiments and found that the SegFormer model exhibited better learning and extraction ability for features contained in wetlands, with various metrics, such as mIoU (0.81), mAcc (0.87), aAcc (0.94), mPrecision (0.901), mRecall (0.876) and mF-score (0.887), being higher than those of other models. We compared the prediction images of various deep learning methods for the key protected area of Yancheng Binhai Wetland and found that SegFormer's prediction results were superior to those of other methods in terms of accuracy and edge clarity, effectively segmenting the boundary in the complex feature environment and extracting small-scale features under influence more accurately. We also studied the common class imbalance in wetlands and concluded that the loss function using the combination of CE_Loss and FocalLoss can improve the classification accuracy of difficult classes and enhance the performance and generalization ability of the model. Finally, we extracted the class proportion information of the Yancheng coastal wetland, which provides technical support and reference value for subsequent wetland value research and research ideas for related coastal wetland information extraction research in China.

Author Contributions: X.L. was accountable for formulating the programs and composing the primary manuscript. Y.C. and G.C. aided in the gathering and analysis of data. W.C. assisted in data collection and produced the figures. R.C. supported in processing remote sensing imagery and generating labels. D.G. played a crucial role in designing and drafting the project, lending vital assistance in manuscript revising. Y.Z. furnished guidance and priceless perspectives on data sources and classification methods. Y.W. charted the project, edited the manuscript and provided constructive feedback throughout the process. All authors have read and agreed to the published version of the manuscript.

Funding: This project is sponsored by the the Jiangsu Forestry Science & Technology Innovation and Extension Project (Project No: LYKJ[2022]02) and Provincial Student Innovation Training Program (Project No: 202310298113Y).

Data Availability Statement: The data are not publicly available due to the confidentiality of the research projects.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bansal, S.; Katyal, D.; Garg, J.K. A novel strategy for wetland area extraction using multispectral MODIS data. *Remote Sens. Environ.* **2017**, *200*, 183–205. [CrossRef]
2. LoweDavid, G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
3. Judah, A.; Hu, B. An Advanced Data Fusion Method to Improve Wetland Classification Using Multi-Source Remotely Sensed Data. *Sensors* **2022**, *22*, 8942. [CrossRef] [PubMed]
4. Suir, G.M.; Jackson, S.; Saltus, C.; Reif, M.K. Multi-Temporal Trend Analysis of Coastal Vegetation Using Metrics Derived from Hyperspectral and LiDAR Data. *Remote Sens.* **2023**, *15*, 2098. [CrossRef]
5. Amani, M.; Brisco, B.; Mahdavi, S.; Ghorbanian, A.; Moghimi, A.; DeLancey, E.R.; Merchant, M.A.; Jahncke, R.; Fedorchuk, L.; Mui, A.; et al. Evaluation of the Landsat-Based Canadian Wetland Inventory Map Using Multiple Sources: Challenges of Large-Scale Wetland Classification Using Remote Sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 32–52. [CrossRef]
6. Li, A.; Song, K.; Chen, S.; Mu, Y.; Xu, Z.; Zeng, Q. Mapping African wetlands for 2020 using multiple spectral, geo-ecological features and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 252–268. [CrossRef]
7. Gxokwe, S.; Dube, T.; Mazvimavi, D. Leveraging Google Earth Engine platform to characterize and map small seasonal wetlands in the semi-arid environments of South Africa. *Sci. Total Environ.* **2021**, *803*, 150139. [CrossRef]
8. Maulik, U.; Chakraborty, D. Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *Isprs J. Photogramm. Remote Sens.* **2013**, *77*, 66–78. [CrossRef]
9. Ahmed, K.R.; Akter, S.; Marandi, A.; Schüth, C. A simple and robust wetland classification approach by using optical indices, unsupervised and supervised machine learning algorithms. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100569. [CrossRef]
10. Fu, B.; Wang, Y.; Campbell, A.D.; Li, Y.; Zhang, B.; Yin, S.; Xing, Z.; Jin, X. Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecol. Indic.* **2017**, *73*, 105–117. [CrossRef]
11. Gonzalez-Perez, A.; Abd-Elrahman, A.H.; Wilkinson, B.E.; Johnson, D.J.; Carthy, R.R. Deep and Machine Learning Image Classification of Coastal Wetlands Using Unpiloted Aircraft System Multispectral Images and Lidar Datasets. *Remote Sens.* **2022**, *14*, 3937. [CrossRef]
12. Munizaga, J.; García, M.; Ureta, F.; Novoa, V.; Rojas, O.; Rojas, C. Mapping Coastal Wetlands Using Satellite Imagery and Machine Learning in a Highly Urbanized Landscape. *Sustainability* **2022**, *14*, 5700. [CrossRef]
13. Prentice, R.M.; Peciña, M.V.; Ward, R.D.; Bergamo, T.F.; Joyce, C.; Sepp, K. Machine Learning Classification and Accuracy Assessment from High-Resolution Images of Coastal Wetlands. *Remote Sens.* **2021**, *13*, 3669. [CrossRef]
14. Xing, L.; Wang, H.; Fan, W.; Chen, C.; Li, T.; Wang, G.; Zhai, H. Optimal Features Selection for Wetlands Classification Using Landsat Time Series. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 8385–8388.
15. Zou, Q.; Chen, C. Transferring Deep Belief Networks for the Classification of LANDSAT8 Remote Sensing Imagery. *J. Phys. Conf. Ser.* **2020**, *1544*, 012106. [CrossRef]
16. Tian, S.; Zhang, X.; Tian, J.; Sun, Q. Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sens.* **2016**, *8*, 954. [CrossRef]
17. Ruiz, L.F.C.; Guasselli, L.A.; Simioni, J.P.D.; Belloli, T.F.; Fernandes, P.C.B. Object-based classification of vegetation species in a subtropical wetland using Sentinel-1 and Sentinel-2A images. *Sci. Remote Sens.* **2021**, *3*, 100017. [CrossRef]
18. Liu, H.; Jiang, Q.; Ma, Y.; Yang, Q.; Shi, P.; Zhang, S.; Tan, Y.; Xi, J.; Zhang, Y.; Liu, B.; et al. Object-Based Multigrained Cascade Forest Method for Wetland Classification Using Sentinel-2 and Radarsat-2 Imagery. *Water* **2022**, *14*, 82. [CrossRef]

19. Zhao, Y.; Mao, D.; Zhang, D.; Wang, Z.; Du, B.; Yan, H.; Qiu, Z.; Feng, K.; Wang, J.; Jia, M. Mapping Phragmites australis Aboveground Biomass in the Momoge Wetland Ramsar Site Based on Sentinel-1/2 Images. *Remote Sens.* **2022**, *14*, 694. [CrossRef]
20. Yang, H.; Liu, X.; Chen, Q.; Cao, Y.B. Mapping Dongting Lake Wetland Utilizing Time Series Similarity, Statistical Texture, and Superpixels With Sentinel-1 SAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8235–8244. [CrossRef]
21. Garba, S.I.; Ebmeier, S.K.; Bastin, J.F.; Mollicone, D.; Holden, J. The Detection of Wetlands And Wetland Fragmentation Using Sentinel 1 And 2 Imagery: The Example of Southern Nigeria. *Res. Sq.* **2021**, preprint.
22. Hosseiny, B.; Mahdianpari, M.; Brisco, B.; Mohammadimanesh, F.; Salehi, B. WetNet: A Spatial-Temporal Ensemble Deep Learning Model for Wetland Classification Using Sentinel-1 and Sentinel-2. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4406014. [CrossRef]
23. Jamali, A.; Mahdianpari, M. Swin Transformer and Deep Convolutional Neural Networks for Coastal Wetland Classification Using Sentinel-1, Sentinel-2, and LiDAR Data. *Remote Sens.* **2022**, *14*, 359. [CrossRef]
24. Jamali, A.; Mohammadimanesh, F.; Mahdianpari, M. Wetland Classification with Swin Transformer Using Sentinel-1 and Sentinel-2 Data. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 6213–6216.
25. Jamali, A.; Mahdianpari, M.; Brisco, B.; Mao, D.; Salehi, B.; Mohammadimanesh, F. 3DUNetGSFormer: A deep learning pipeline for complex wetland mapping using generative adversarial networks and Swin transformer. *Ecol. Inform.* **2022**, *72*, 101904. [CrossRef]
26. He, X.; Zhou, Y.; Zhao, J.; Zhang, D.; Yao, R.; Xue, Y. Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4408715. [CrossRef]
27. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple Multiscale UNet for Change Detection With Heterogeneous Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2504905. [CrossRef]
28. Sharma, A.; Liu, X.; Yang, X.; Shi, D. A patch-based convolutional neural network for remote sensing image classification. *Neural Netw.* **2017**, *95*, 19–28. [CrossRef] [PubMed]
29. Li, X.; Xu, F.; Lyu, X.; Gao, H.; Tong, Y.; Cai, S.; Li, S.; Liu, D. Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images. *Int. J. Remote Sens.* **2021**, *42*, 3583–3610. [CrossRef]
30. Raza, A.; Huo, H.; Fang, T. EUNet-CD: Efficient UNet++ for Change Detection of Very High-Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3510805. [CrossRef]
31. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [CrossRef]
32. Bazi, Y.; Bashmal, L.; Al Rahhal, M.M.; Dayil, R.A.; Ajlan, N.A. Vision Transformers for Remote Sensing Image Classification. *Remote Sens.* **2021**, *13*, 516. [CrossRef]
33. He, J.; Zhao, L.; Yang, H.; Zhang, M.; Li, W. HSI-BERT: Hyperspectral Image Classification Using the Bidirectional Encoder Representation From Transformers. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 165–178. [CrossRef]
34. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.J.; Chanussot, J. SpectralFormer: Rethinking Hyperspectral Image Classification With Transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5518615. [CrossRef]
35. Mohammadimanesh, F.; Salehi, B.; Mahdianpari, M.; Brisco, B.; Motagh, M. Multi-temporal, multi-frequency, and multi-polarization coherence and SAR backscatter analysis of wetlands. *Isprs J. Photogramm. Remote Sens.* **2018**, *142*, 78–93. [CrossRef]
36. Brendel, W.; Bethge, M. Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet. *arXiv* **2019**, arXiv:1904.00760.
37. Azulay, A.; Weiss, Y. Why do deep convolutional networks generalize so poorly to small image transformations? *J. Mach. Learn. Res.* **2018**, *20*, 181–184.
38. Ilyas, A.; Santurkar, S.; Tsipras, D.; Engstrom, L.; Tran, B.; Madry, A. Adversarial Examples Are Not Bugs, They Are Features. *arXiv* **2019**, arXiv:1905.02175.
39. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv* **2018**, arXiv:1811.12231.
40. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Álvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *arXiv* **2021**, arXiv:2105.15203.
41. Lane, C.R.; Liu, H.; Autrey, B.C.; Anenkhonov, O.A.; Chepinoga, V.V.; Wu, Q. Improved Wetland Classification Using Eight-Band High Resolution Satellite Imagery and a Hybrid Approach. *Remote Sens.* **2014**, *6*, 12187–12216. [CrossRef]
42. Amani, M.; Mahdavi, S.; Bérard, O. Supervised wetland classification using high spatial resolution optical, SAR, and LiDAR imagery. *J. Appl. Remote Sens.* **2020**, *14*, 024502. [CrossRef]
43. Amani, M.; Foroughnia, F.; Moghimi, A.; Mahdavi, S. 3D Habitat Mapping Using High-Resolution Optical Satellite and Lidar Data. In Proceedings of the 2022 10th International Conference on Agro-geoinformatics (Agro-Geoinformatics), Quebec City, QC, Canada, 11–14 July 2022; pp. 1–5.
44. Kaplan, G.; Avdan, U. Evaluating the utilization of the red edge and radar bands from sentinel sensors for wetland classification. *Catena* **2019**, *178*, 109–119. [CrossRef]
45. Wu, N.; Shi, R.; Zhuo, W.; Zhang, C.; Zhou, B.; Xia, Z.; Tao, Z.; Gao, W.; Tian, B. A Classification of Tidal Flat Wetland Vegetation Combining Phenological Features with Google Earth Engine. *Remote Sens.* **2021**, *13*, 443. [CrossRef]

46. Li, X.; Xu, F.; Liu, F.; Xia, R.; Tong, Y.; Li, L.; Xu, Z.; Lyu, X. Hybridizing Euclidean and Hyperbolic Similarities for Attentively Refining Representations in Semantic Segmentation of Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5003605. [CrossRef]
47. Araya-López, R.A.; Lopatin, J.; Fassnacht, F.E.; Hernández, H.J. Monitoring Andean high altitude wetlands in central Chile with seasonal optical data: A comparison between Worldview-2 and Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 213–224. [CrossRef]
48. Liu, J.; Feng, Q.; Gong, J.; Zhou, J.; Li, Y. Land-cover classification of the Yellow River Delta wetland based on multiple end-member spectral mixture analysis and a Random Forest classifier. *Int. J. Remote Sens.* **2016**, *37*, 1845–1867. [CrossRef]
49. Jiao, L.; Sun, W.; Yang, G.; Ren, G.; Liu, Y. A Hierarchical Classification Framework of Satellite Multispectral/Hyperspectral Images for Mapping Coastal Wetlands. *Remote Sens.* **2019**, *11*, 2238. [CrossRef]
50. Li, X.; Xu, F.; Liu, F.; Lyu, X.; Tong, Y.; Xu, Z.; Zhou, J. A Synergistical Attention Model for Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5400916. [CrossRef]
51. Li, X.; Xu, F.; Xia, R.; Li, T.; Chen, Z.; Wang, X.; Xu, Z.; Lyu, X. Encoding Contextual Information by Interlacing Transformer and Convolution for Remote Sensing Imagery Semantic Segmentation. *Remote Sens.* **2022**, *14*, 4065. [CrossRef]