remote sensing

MDPI

Article

SiamHRnet-OCR: A Novel Deforestation Detection Model with High-Resolution Imagery and Deep Learning

Zhipan Wang [1,2], Di Liu [1,2], Xiang Liao [3], Weihua Pu [4], Zhongwu Wang [5] and Qingling Zhang [1,2,*]

[1] Shenzhen Key Laboratory of Intelligent Microsatellite Constellation, Shenzhen Campus of Sun Yat-sen University, No. 66, Gongchang Road, Guangming District, Shenzhen 518107, China
[2] School of Aeronautics and Astronautics, Shenzhen Campus of Sun Yat-sen University, No. 66, Gongchang Road, Guangming District, Shenzhen 518107, China
[3] Chongqing Pioneer Satellite Technology Co., LTD., Chongqing 401420, China
[4] Shenzhen Aerospace Dongfanghong Satellite Co., LTD., Shenzhen 518061, China
[5] Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, Beijing 100048, China
* Correspondence: zhangqling@mail.sysu.edu.cn

Abstract: Forests play a critical role in global carbon cycling, and continuous forest destruction together with other factors has accelerated global warming. Furthermore, continued decline of forest areas will critically hinder the accomplishment of carbon neutrality goals. Although the geographical location of deforestation can now be rapidly and accurately detected with remote sensing technology, current forest change products are still not fine-grained, especially from the perspective of carbon trading. Here, we used a deep learning method to detect deforestation in large regions based on 2 m high-resolution optical remote sensing images. Firstly, we proposed a new deforestation detection dataset, which was generated from 11 provincial regions in the Yangtze River Economic Zone of China, containing a total number of 8330 samples (the size of each sample being 512 × 512 pixels). Then, a new deforestation detection model, SiamHRnet-OCR, was designed, based on this dataset. Compared with other deep learning models, SiamHRnet-OCR achieves better results in terms of precision, F1-score, and OA indicator: 0.6482, 0.6892, and 0.9898, respectively. Finally, two large-scale scenarios of deforestation experiments in Southern China were further tested; the deforestation detection results demonstrate that SiamHRnet-OCR can not only detect deforestation effectively but also capture the accurate boundary of the changing area.

Keywords: carbon neutrality; deep learning; deforestation detection; large-scale applications; change detection

check for updates

Citation: Wang, Z.; Liu, D.; Liao, X.; Pu, W.; Wang, Z.; Zhang, Q. SiamHRnet-OCR: A Novel Deforestation Detection Model with High-Resolution Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, 463. https://doi.org/10.3390/rs15020463

Academic Editor: Michael K Ng

Received: 21 November 2022
Revised: 2 January 2023
Accepted: 10 January 2023
Published: 12 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In 2020, the United Nations released the Global Forest Resources Assessment report, which stated that, since 1990, a staggering 178 million hectares of forest have been lost worldwide, either legally or illegally [1]. Continued forest loss will have a major impact on the global climate balance and hinder the achievement of the set goal of carbon neutrality [2,3]. Developed countries around the world, such as Japan, the EU, the UK, and Canada, have set their own carbon neutrality deadlines. Developing countries, such as China, are also striving to reach the goal of carbon neutrality. Emission trading is now a common practice to help cap carbon emissions in many countries [4]. Therefore, accurate and timely information on forest change is essential for accurate carbon accounting and carbon neutrality.

Deforestation is mainly caused by forest change, due to natural factors, or human activities [5]. Natural factors include diseases of trees, forest fires, parasites, and extreme weather such as floods or hurricanes [6]. Human activities also play an important role in deforestation [7]—for example, farmland reclamation, infrastructure construction, mining activities, and urbanization [8]. Remote sensing imagery has some important advantages,

such as the free-use policy and longhistorical archives, which make it has become the main data source to monitor forest change on the earth [9–15]. For example, the Landsat images are the most widely used data source to monitor deforestation so far because of their long historical archives (over 40 years) and especially their open-source and free-use policy in place since 2008 [16]. Based on the Landsat images, some excellent forest change research works have been carried out in tropical regions [17], temperate regions [11], and even at the global scale [4], and many classic algorithms have been proposed [17–21]. For instance, the CCDC algorithm (Continuous Change Detection and Classification) [22], can detect landcover change by modeling the change of any pixel in the same area over a long period. However, the limitation of CCDC is that the calculation cost is heavy and slow. A further improved method, namely S-CCD (Stochastic Continuous Change Detection) [23], was proposed to solve this problem. S-CCD relieves the computation burden by treating seasonal forest changes as stochastic processes and then introducing a mathematical tool called the "State Space Model" to detect changes. All the existing proposed methods have excellent performance with 30 m or 10 m median spatial resolution imagery. However, after analyzing the existing forest change products, researchers pointed out that forest change areas counted by the existing methods show large uncertainty, the relatively coarse spatial resolution of the remote sensing image being the main factor in this discrepancy [24].

Recently, using deep learning methods to monitor deforestation on median resolution imagery, such as Landsat8 [25] or Sentinel2A/2B [26] images, has attracted much attention. For example, [12] used a ResUnet model to monitor deforestation detection with Landsat8 and Sentinel2A/B, and demonstrated that the deep learning method outperforms traditional machine learning methods such as the Random Forest classifier. Additionally, high-resolution imagery, such as Planet (3.7 m) [27] and Komsat-3 (0.7 m) [28], has also been used to monitor deforestation with deep learning methods, with good detection accuracy. However, there is still a lack of high-quality deforestation training datasets for the community to use in training deep learning models. It is well known that a high-quality training dataset is very important for training good deep learning models, but the process of generating a large-size training dataset is very time-consuming and expensive.

Another factor affecting the accuracy of deforestation detection is the structure of deep learning models. Though there are some deforestation detection models that have been proposed, such as Unet [13], DeepLabV3+ [29], improving the model structure still has the potential to improve detection accuracy. For example, the attention Unet [15] achieves better accuracy than Unet or other segmentation models. However, most of the existing deforestation detection models cannot maintain high-resolution semantic features forwarding during the whole training process, which will decrease the detection accuracy on a narrow object or other complex regions [30]. In this manuscript, we proposed a new high-resolution deep deforestation detection network, namely SiamHRnet-OCR, which shows better detection accuracy than existing models. The main advantage of SiamHRnet-OCR is that high-resolution features forwarding is always kept in the whole model layers.

The major contributions of this manuscript are as follows:

(1) A new deforestation training sample dataset was proposed, containing a total of 8330 true color samples (512 × 512 pixels) of 2 m spatial resolution. This dataset was generated by visual interpretation in 11 provinces of China's Yangtze River Economic Zone, and it will be open-sourced to the community to help researchers worldwide to conduct deforestation detection studies.

(2) A new deforestation detection model, namely the SiamHRnet-OCR, can effectively improve detection accuracy, especially for narrow objects or complex regions.

(3) The design principle of SiamHRnet-OCR can provide some new insights for other research fields, for example, road or building change detection.

Related Work

Deforestation detection based on deep learning methods is a hot research topic. Both optical-based and SAR-based deep learning methods have been proposed [13,14]. On

the whole, most of the existing deforestation detection modules are encoder–decoder structures, with the Unet style being the most commonly used structure. For example, the ForestNet [14] was designed with an encoder–decoder structure to classify the drivers of primary forest loss in Indonesia, and the results showed that it outperformed than the Random Forest method [31]. To alleviate the cloud effect on optical remote sensing images, the dense time series Sentinel-1 SAR imagery was used to map forest harvesting in California, USA and Rondonia, Brazil [13], with a simple Unet module. In addition, the Siamese CNN(S-CNN) [32], the Pyramid Feature Extraction Unet model (PFE-UNet) [33], the DeepLabv3+ model [29], and the LSTM and CNN combined model [34] are also used to detect deforestation. In order to extend the feature extraction ability of CNN, the attention module [35] was also investigated and indicates better precision than the pure CNN structure. For example, [15] proposed an attention-based Unet model to detect deforestation, and the results show that the attention Unet has higher accuracy than the Unet, Residual Unet, ResNet50-SegNet, and the FCN32-VGG16 models with Sentinel-2 optical remote sensing images.

Deforestation detection can be defined as a classical change detection task, and it can also be understood as an extension of pixel-level image classification [9]. Therefore, other excellent change detection models designed for building change detection or other domains can provide some new insights for deforestation detection, such as the SiamFCN [36], Unet++ [37], STAnet [38], DTCDSCN [39], ESCNet [40], and SNUNet [41]. In the aforementioned models, the whole feature extraction process contains three main steps: firstly, the backbones module is used to extract multi-scale low-level and high-level semantic features, such as ResNet [42], MobileNet [43], etc. Secondly, multi-scale semantic features are fused by concatenation and skip-connection operation [44–46]. Finally, a loss function is used to optimize the feature extraction direction. In the whole process, a critical question is to design a reasonable deep-learning model architecture; to acquire rich and effective semantic features of objects. However, the downsampling operators in most deep learning models lead to irreversible information loss [47], especially for pixel-level classification in remote sensing imagery. As a result, the accuracy of change detection may be decreased, especially in the boundary or pseudo-change regions. The HRnet (High-resolution network) proposed by [30] has achieved state-of-the-art accuracy in semantic segmentation tasks on naturalistic images. The main advantage of HRnet is that it can capture effective context features of small targets, such as tree trunks and traffic lights, because it always delivers deep semantic features in high resolution during the whole feature extraction process. In remote sensing images, the clarity of objects is mainly determined by the spatial resolution if high-resolution features can be kept in the entire semantic feature extraction process; therefore, even insignificant spectral changes and slight texture changes in remote sensing imagery can in theory be distinguished based on effective high-level semantic features.

2. Study Area

We conducted experiments in two large regions in southern China because a recent study reported many deforestation hotspots there [48]. The bi-temporal images of the two study areas are shown in Figure 1.

The first study area was Hengyang City in Hunan Province, China (chosen as the main study area due to the diverse land cover types and lower urban area proportion in this region). Hengyang city is located in the central-south of China, and the land area is approximately 2621 km^2. The region has a subtropical monsoon climate, and the terrain is mainly hilly. The major forest types are evergreen broad-leaved forest, deciduous broad-leaved forest, and evergreen coniferous forest; the majority of trees are planted forests with few original forest coverages. The growth cycle of planted forests is fast, usually 5–10 years before they can be cut down for making furniture and hand tools. From the relevant public government's statistical data, deforestation frequently occurs in this region.

The second study area is Qujing City in Yunnan Province. Qujing is located in the southwestern China, with a subtropical plateau monsoon climate and a whole area of

approximately 28,900 km^2. The forest coverage type in this region is mainly primary forests. In recent years, China's policies of "Poverty Alleviation" and "Common Prosperity" have increased investment in this region; consequently, more and more infrastructure such as highways and railways have been built here. In addition, to improve the living conditions of the original inhabitants, many cultivated lands were also developed in this region, and as a result deforestation in the region become serious in recent years.

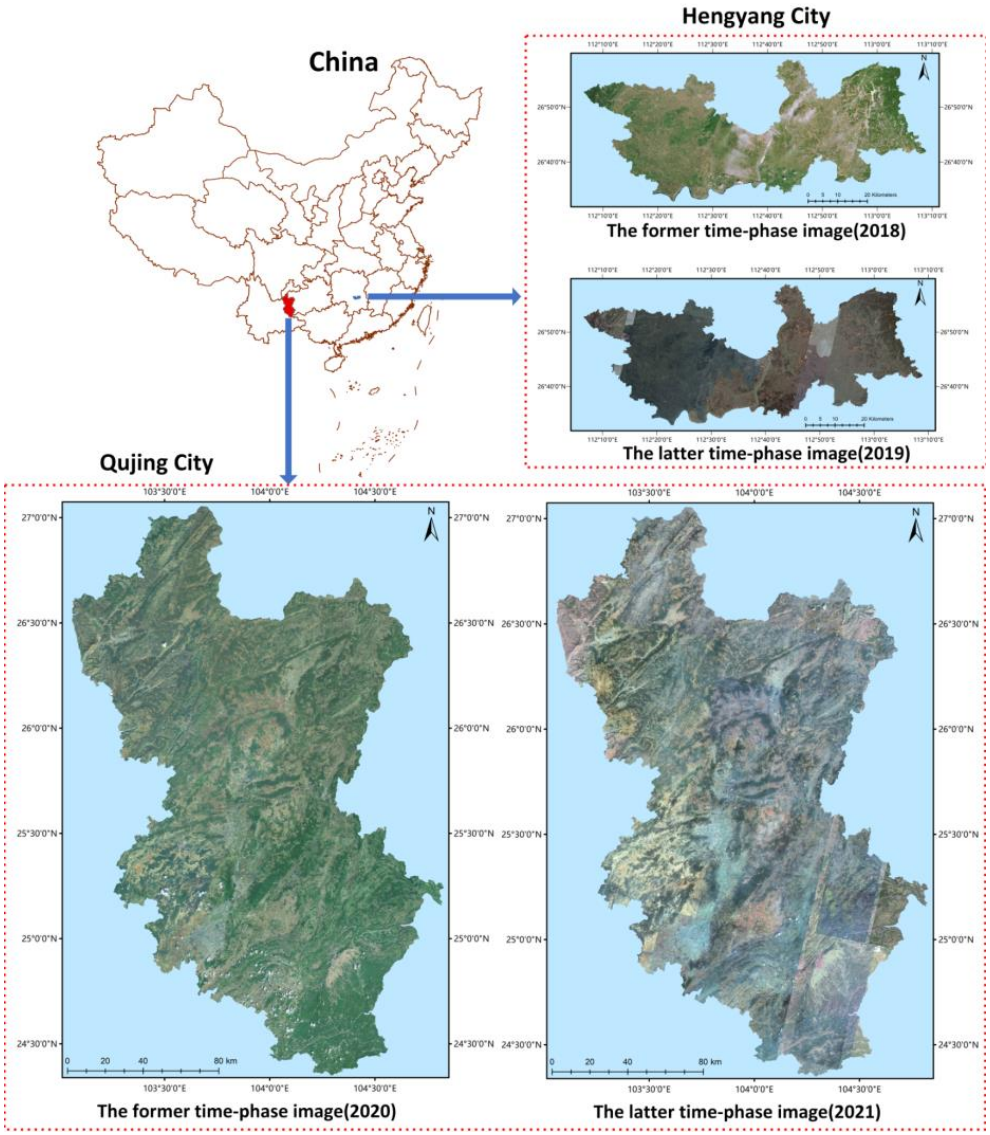


Figure 1. Study areas.

3. Materials and Methods

3.1. Experimental Data

The metadata information, such as the image acquisition time, spatial resolution, data source, and band information of the two study areas, can be seen in Table 1. It is worth noting that the image acquisition time between Hengyang City and Qujing City is different. Furthermore, the image quality of the two study areas in the latter period is not very good; for example, some clouds or multiple image mosaicking footprints can be identified in Figure 1.

Table 1. The metadata information of remote sensing images in the study areas.

Study Area	Acquisition Time		Spatial Resolution (m)	Data Source	Bands
	The Former Period	The Latter Period			
Hengyang City	June, July, and August in 2018	June, July, and August in 2019	2	GF-1, GF-2, ZY-3	R, G, B
Qujing City	June, July, and Augustin in 2020	From June to September in 2021	2	GF-1, GF-2, ZY-3	R, G, B

The bi-temporalremote sensing imageries of Hengyang City and Qujing City are from different sources, including 2 m GF-1, 0.8 m GF-2, and a few 2.1 m ZY-3 images. GF-1, GF-2, and ZY-3 are the optical remote sensing satellites launched by China; the spatial resolution and other specifications can be seen in Table 2. This whole dataset can be downloaded from https://www.cresda.com/zgzywxyyzx/index.html (accessed on 12 September 2021). All of the remote sensing imageries used are provided by the Land Satellite Remote Sensing Application Center of China, and all of the images have been preprocessed by the Pansharpenging and Geometric Correction module based on PCI GXL software.

Table 2. The metadata information of remote sensing imagery.

Name	Launched Year	Spatial Resolution	Swaths	Band
GF-1	2013	2 m	60 km	R, G, B, NIR
GF-2	2014	0.8 m	45 km	R, G, B, NIR
ZY-3	2012	2.1 m	51 km	R, G, B, NIR

In order to keep the spatial resolution consistent, the GF-2 and ZY-3 images are resampled to 2 m with a nearest neighbor resampling method. In Qujing City, the latter period images were taken in summer and autumn, because in this region there is always a high probability of cloud cover in summer, and extending the observation period is the only way to get higher-quality optical remote sensing images. Moreover, the image band of the two study areas only have visible bands.

3.2. Deforestation Detection Sample Datasets

Samples are the most important part in developing deep learning models. To the best of our knowledge, there is still a lack of a high-quality and high-resolution deforestation detection sample datasets so far. In this manuscript, we have developed such a dataset for the first time, generated from 2 m mosaic images produced by China Resources Satellite Data and Application Center. The remote sensing imagery used in this study were obtained from https://data.cresda.cn/#/2dMap (accessed on 1 July 2021). The meta-information about this dataset can be seen in Table 3 (the 1 m GF-2 data were resampled to 2 m with the nearest neighbor sampling method).

Table 3. The meta-information of the high-resolution deforestation detection dataset.

Acquisition Regions	Acquisition Time		Spatial Resolution (m)	Data Source	Bands	Sample Size	Number
	The Former Period	The Latter Period					
The Yangtze River Economic Zone of China	From May to September in 2019	From May to October in 2020	2	GF-1, GF-2, ZY-3, GF-6	R, G, B	512 × 512 Pixels	8330

The mosaicing image, which was used to generate training samples, was radiometrically and geometrically corrected, and it only contained visible bands. To alleviate the cloud effect and to improve data availability, most of the source remote sensing images were acquired in the summer and autumn of the whole year. The most important advantage of this deforestation detection dataset is that it was generated from a very large region (the Yangtze River Economic Zone of China), including 11 provinces in China: Shanghai, Anhui,

Zhejiang, Jiangxi, Jiangsu, Hunan, Hubei, Guizhou, Yunnan, Chongqing, Sichuan. Most of these provinces are mainly located in the subtropical region, except for a few cities in Yunnan Province that are in the tropical region, such as the Lincang city and the Tengchong city. The Yangtze River Economic Zone hosts the majority of China's population, and due to the rapid development of China's economy, deforestation has become more serious in this region in recent years. Consequently, it is representative enough to construct a deforestation detection sample dataset in such a large region with varying terrains and landscapes.

The production process of the deforestation detection sample datasets was carried out in a commercial web-based system (a commercial webGIS system developed for internal use), and the bi-temporal images were published as an image map through the WMTS service. A staff member who has administrator access can work in this system, outlining deforestation change regions and checking the sample quality, The whole process contains four main steps (Figure 2): (1) the whole area of China Yangtze River Economic Belt was divided into many 2 km × 2 km grids; (2) a staff member used the swipe tool to compare differences between the two time-phase images (2019 and 2020) in a local area (for example, 100 km^2) and then outline the boundaries of deforestation in each grid; (3) fixed-size samples (512 × 512 pixels) were output according to the requirements of chunking; (4) finally a manual quality check was performed to ensure that the whole China Yangtze River Economic Zone was covered by deforestation samples with accurate boundaries. It is worth mentioning that our definition of forest refers to land covered by trees, including deciduous broadleaf forest, evergreen broadleaf forest, deciduous coniferous forest, evergreen coniferous forest, mixed forest, and sparse forest with 10–30% canopy cover [49].

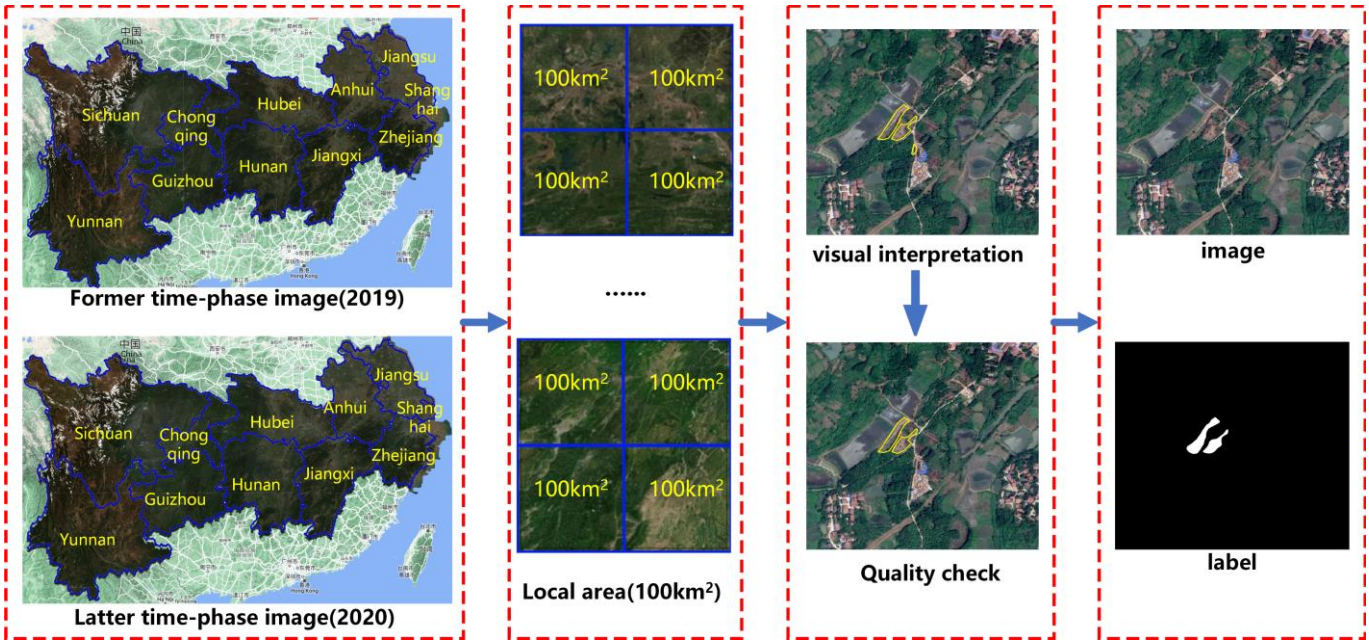


Figure 2. The flowchart of sample dataset generation.

A few samples from our deforestation detection datasets are shown in Figure 3. We can see that this deforestation detection dataset contains various types of deforestation, including forest destruction due to human-induced development and construction (such as road and building construction), natural fire, and other uncontrollable factors. To facilitate researchers to conduct deforestation detection research, it can be downloaded from (https://drive.google.com/drive/folders/1LtgNjmDePqD6mJkAepO0GUEi1uMwPRUQ?usp=sharing (accessed on 15 December 2022)).

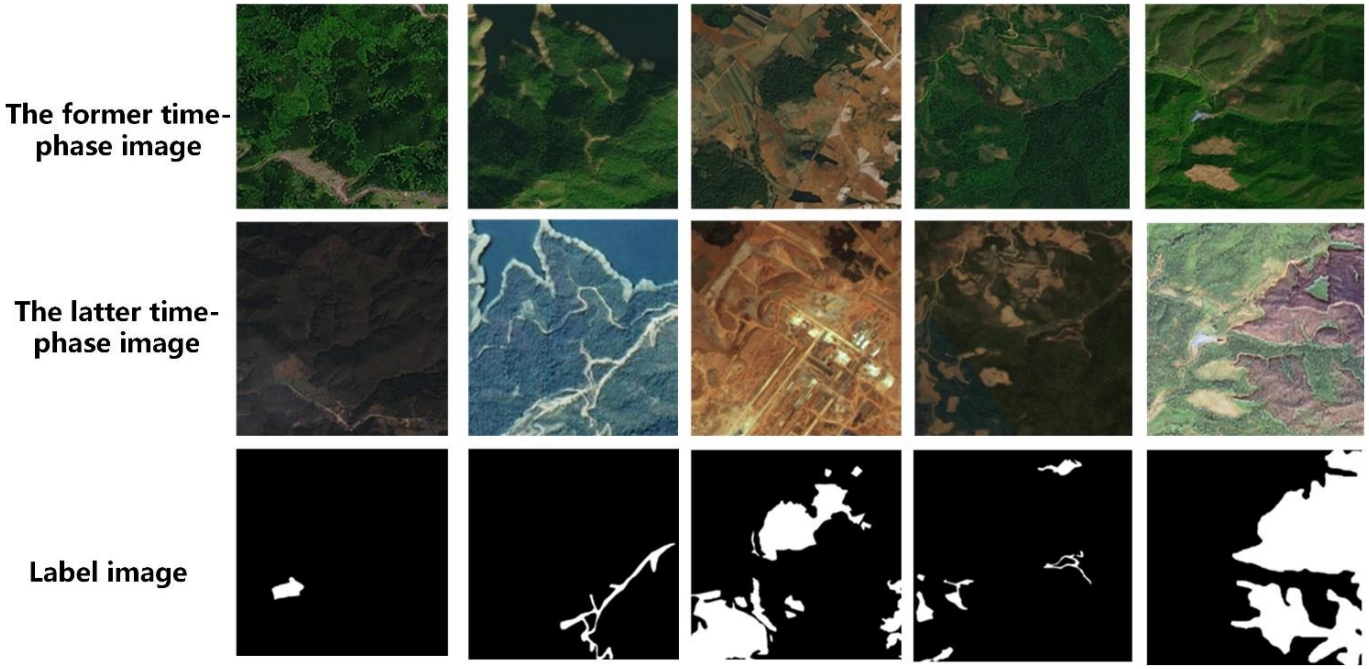


Figure 3. Deforestation detection dataset.

To reasonably evaluate the scientificity and rationality of the deforestation detection dataset, we select some excellent change detection datasets for comparison, for example, the SYSU-CD dataset [50], which was generated for multi-class change detection, including roads, buildings, cropland, and water. The Google dataset [51] also provided a sample dataset for waters, roads, bare land, and building change detection. Furthermore, in the building change detection field, a few carefully generated datasets have been proposed, such as the WHU building datasets [52], the LEVIR-CD dataset [38], and the S2Looking dataset [53]. Detailed information about different datasets can be seen in Table 4.

Table 4. Change detection datasets comparison.

Dataset	Types of Land Surface Changes	Data Description
SYSU-CD [50]	roads, buildings, cropland, water	20,000 pairs of 0.5-m aerial images (The size of each image is 256 × 256 pixels, which were clipped from 5 large images. These large images were acquired in Beijing, Chengdu, Shenzhen, Wuhan, Chongqing, and Xian, China.
LEVIR-CD [38]	Buildings	637 pairs of 0.5 m google earth images (The size of each image is 1024 × 1024 pixels), Collected from Texas, American.
WHU building datasets [52]	Buildings	About 20,000 pairs of 2.7 m google earth images (The size of each image is 512 × 512 pixels), and 8000 pairs of 0.1 m aerial images (The size of each image is 512 × 512 pixels), Collected from Wuhan, Taiwan, New York, and other places (1000 km^2 in total).
Google datasets [51]	waters, roads, farmland, bare land, buildings	9 VHR image pairs of 0.55 m, the image size ranging from 1006 × 1168 pixels to 4936 × 5224 pixels, and this dataset was Collected from Guangzhou city, China.
S2Looking [53]	buildings	5000 pairs of 0.5~0.8 m images (the size of each image is 1024 × 1024 pixels).
This study	deforestation	8330 pairs of 2 m RGB images (the size of each image is 512 × 512 pixels, and this dataset was collected from Hunan province, Hubei province, Guizhou province, Jiangsu Province, and other places (11 provinces in total, 2,052,300 km^2), China.

In the above table, the existing open-sourced change detection datasets include different land cover change types; however, all of their training samples are generated from a single large image by slicing it into small image blocks (for example, a 5000 × 5000 pixels image are randomly cropped into small blocks with 512 × 512 pixels). The workflow of

sample production is simple, but the model generalization ability of the trained model is limited [38]. Unlike these existing training sample construction methods, we produce training samples discretely over a very large region (11 provinces of China), which can effectively improve model performance in real-world scenarios.

As we know, a high-quality sample dataset is the most important component of remote sensing interpretation tasks. Although the number of our deforestation detection dataset is not obvious, the acquisition region of this dataset is much larger than other existing open-sourced datasets. Another advantage of this dataset is that each training sample contained deforestation pixels (they can also be defined as positive pixels [9]), which will increase the focus of the deep learning model on deforestation areas. In this deforestation detection dataset, the ratio between changed pixels (positive samples) and unchanged pixels (negative samples) is about 1:11.8. To train an effective deforestation detection model, the whole deforestation detection dataset was split into three sub-datasets: train dataset, validation dataset, and test dataset, as shown in Table 5.

Table 5. Deforestation detection dataset split.

Dataset	The Number of Sample Images	Usage
Train	6730	Train change detection model
Validation	800	Selecting the best model
Test	800	Accuracy assessment

3.3. Proposed SiamHRnet-OCR Change Detection Model

There are two major limitations to the existing deep learning-based change detection methods [47]. First, existing models pay much attention to increasing the receptive field to acquire richer contextual semantic features, such as the ASPP (Atrous Spatial Pyramid Pooling) [44] and the PPM (Pyramid Pooling Module) [45], etc., but whether there is a strong correlation between the contextual features acquired by these modules and object types were rarely studied. Second, generally speaking, remote sensing images and naturalistic images (real-life situations or scenarios) are different, and the sharpness of objects varying greatly at different spatial resolutions, and thus simply applying the semantic segmentation methods designed for naturalistic images to RS images [51,54] needs further investigation, especially for deforestation detection, because the context features of some changed regions are complex and the shape are also more irregular [15].

The HRnet semantic segmentation model [30] is a high-resolution semantic segmentation network, enabling high-resolution feature forwarding on the whole model layers. Therefore, HRnet is well suited for spatial-scale sensitive tasks, such as road extraction, change detection, etc. The subsequently developed OCR (Object Context Representation) module [47] mainly focuses on the feature correlation of the same object, constructing a finer feature representation by modeling the global segmentation of the initial object; the final segmentation result confirms its effectiveness. Inspired by HRnet [30] andOCR [47], we proposed a new deforestation detection model, namely the SiamHRnet-OCR (see Figure 4). The main architecture of the model is a Siamese structure [38], and the core idea of SiamHRnet-OCR is that two weight-shared HRnets are used as the backbone to acquire rich semantic features at different levels before an OCR module is used to refine the deep semantic features to further focus on the change regions [47].

The whole architecture of the SiamHRnet-OCR can be divided into four parts. the deep feature extraction module, the deep feature fusion module, the OCR refine module, and the change result optimization module. As shown in Figure 4, the most important characteristic of the SiamHRnet-OCR is that high-resolution features are always maintained during the whole training process. In the SiamHRnet-OCR model, the deep feature fusion module is used to locate the differences between the deep semantic features of the bi-temporal images. Compared with ASPP [44], PPM [45] or other context feature extraction modules, the most important advantage of the OCR refine module is that it uses object features for context modeling, allowing it to extracted richer and finer features than other modules. The change

result optimization module uses a loss function to optimize the feature learning direction, which is the last important part of the SiamHRnet-OCR model.

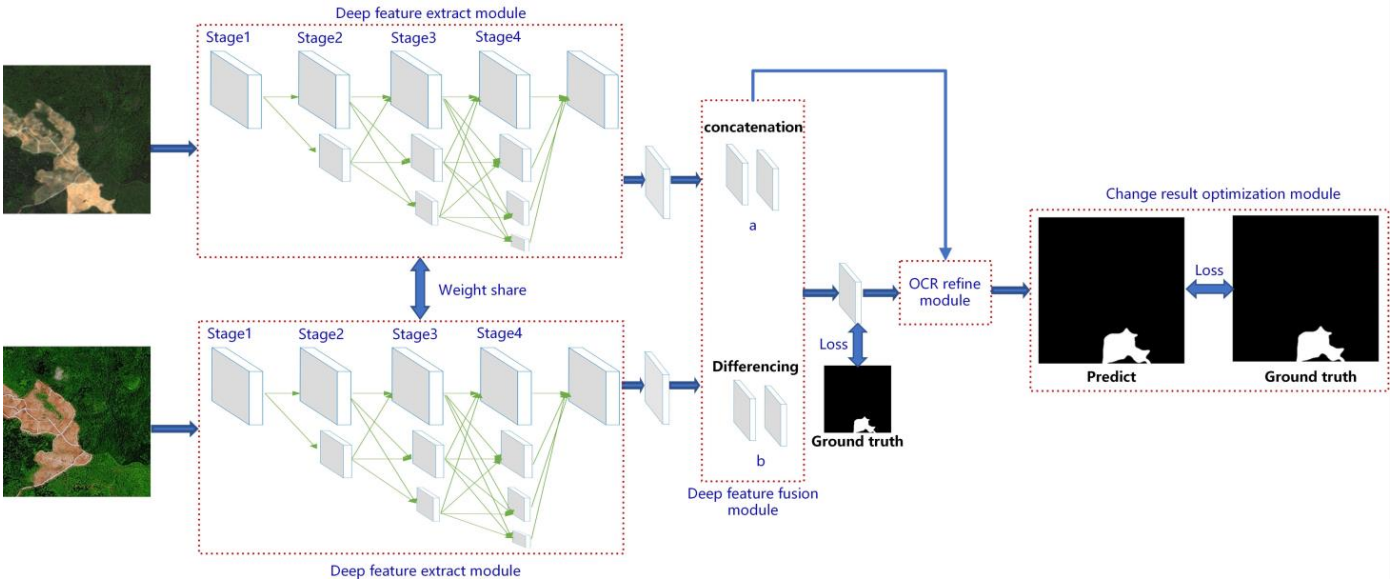
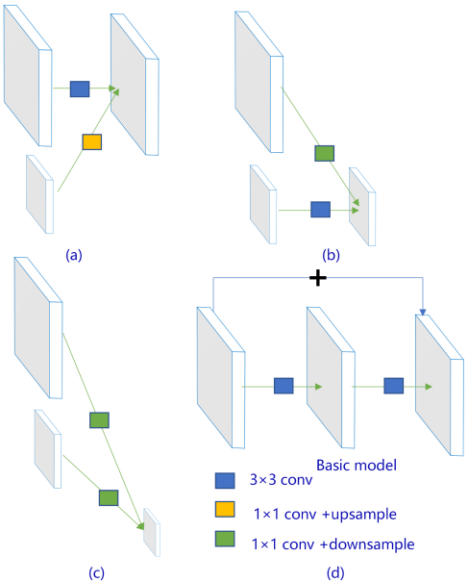


Figure 4. The architecture of the SiamHRnet-OCR.

3.3.1. The Deep Feature Extraction Module

We use a Siamese network architecture to extract the deep semantic features of bi-temporal images [50,54], and the weight parameters in different stages are shared. The module is divided into four stages, and each stage uses the residual network connectivity to deepen the model depth [42]. The architecture of the residual module and the parameters of each layer can be seen in Figure 5. The reason we use residual network connectivity [47,55,56] is that a residual network can relieve the gradient disappearance problem as the network is deepening [57].Moreover, the residual network can also obtain rich semantic features. In the whole "Deep feature extraction module", high-resolution semantic features are always retained, which is extremely important to improve model's performance in boundary regions or small object detection accuracy [47].



layer name	output size	Channel	Operator
input	512×512	3	3×348 stride 4
Stage1	128×128	48	[basic model]×2
Stage2	128×128	48	[basic model]×4
	64×64	96	[basic model]×4
Stage3	128×128	48	[basic model]×4
	64×64	96	[basic model]×4
	32×32	192	[basic model]×4
Stage4	128×128	48	[basic model]×4
	64×64	96	[basic model]×4
	32×32	192	[basic model]×4
	16×16	384	[basic model]×4
output	128×128	360	3×3360

(e) Parameter at each stage

Figure 5. (**a**–**c**) Feature forward from stage 2 to stage 3. (**d**) The basic model of stages 1–4. (**e**) Parameter at each stage.

In Figure 5, during the feature extraction process, we use a 3 × 3 convolution for feature forwarding and then use an additional operator to fuse semantic features with different resolutions [43]. There are two cases for feature forwarding with different resolutions. In the first case, for features from low-resolution to high-resolution (Figure 5a), a 1 × 1 convolution and an upsampling operator are used for feature transferring. In the second case, feature transferring from a high-resolution to a low-resolution, we use a 1 × 1 convolution and a downsampling operator (see in Figure 5b,c) to keep feature size consistent. In order to speed up computational efficiency, we used only the nearest-neighbor interpolation to implement downsampling and upsampling; the detailed parameter information from the input layer to the output layer can be seen in Figure 5e.

3.3.2. The Deep Feature Fusion Module

After the high-level semantic feature is extracted by the backbone, we use two feature fusion modules for the next step of feature learning, namely "differencing" and "concatenation" (Figure 4). In the "differencing" deep feature fusion method we use feature difference to fuse the bi-temporal deep semantic change features, while in the "concatenation" deep feature fusion method we fuse the bi-temporal deep semantic change features with a stack operation. In this manuscript, the "differencing" operator is the default way to fuse the high-level semantic features, because the existing study [38] indicates that the "differencing" method achieves better precision than the "concatenation" method; the experiment in Section 5.2 also confirms this point.

3.3.3. The OCR Refine Module

The main idea of the OCR refine module is that it can characterize each pixel by exploiting the corresponding object representation and aggregating high-level semantic features to the object itself [47]. Compared with direct segmentation of the HRnet, the segmentation accuracy of the OCR module can be improved to a certain extent, especially in cases with complex backgrounds [47]. The OCR refine module first makes an initial change judgment to obtain the approximate change area, and then aggregates these high-level features to obtain finer change detection results. The core operator in the OCR refine module is the multiple matrix multiplication operation, which can be generalized as an attention mechanism. The recent SegNext [58] also demonstrates that matrix multiplication is simple but useful for dense pixel-level segmentation tasks. The computational flow of this module is shown in Figure 6.

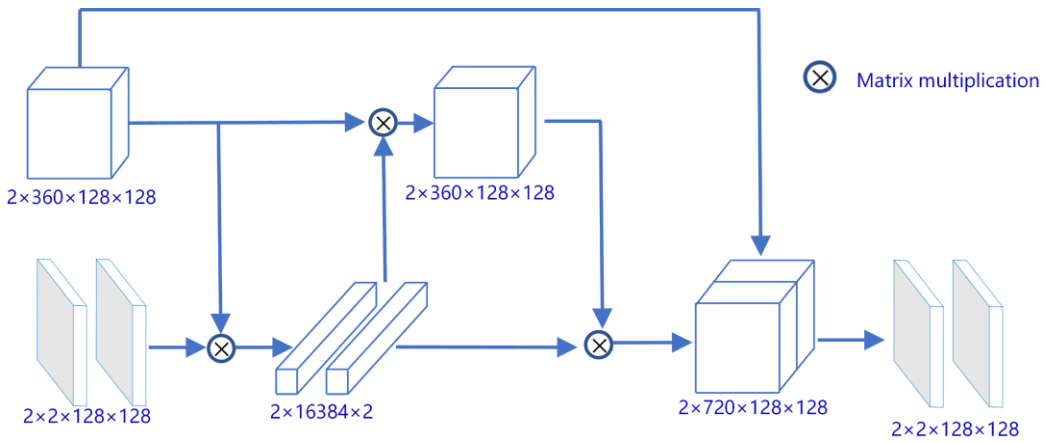


Figure 6. OCR refine module.

3.3.4. The Change Result Optimization Module

The loss function is important for model training, and it also significantly affects the accuracy of change detection results. Currently, the commonly used loss function is cross-entropy [38]. For a pixel-level change detection task, there are often relatively few

changed pixels in bi-temporal images. Consequently, if we directly use the cross-entropy loss function, it will cause the deep learning model to pay more attention to these stable unchanged regions increasing the omission alarms of the final detection result. Therefore, we construct a new loss function—*MWCE* (Modified Weighted Cross-Entropy), which is sensitive to an unbalanced class. The formula is as follows:

$$MWCE = -\left\{ \frac{\sum\limits_{i \in Z_1} \log P(y_i = 1)}{N_{\mathrm{p}}} + \frac{\sum\limits_{i \in Z_2} \log(1 - P(y_i = 0))}{N_n} \right\} \tag{1}$$

where, $P(y_i = 1)$ denotes the probability value of a positive sample pixel i through a softmax function, $P(y_i = 0)$ the probability value of a negative sample pixel i through a softmax function, Z_1 the set of positive sample pixels, Z_2 the set of negative sample pixels, N_{p} and N_{n} the number of pixels in a labeled image for positive and negative samples, respectively.

3.4. Accuracy Assessment

We use Recall, Precision, F1, and OA (Over Accuracy) [37] for accuracy assessment. The Recall indicator reflects the omission error of positive samples; the higher the value the lower the omission of positive samples. The Precision indicator reflects the commission error of positive samples; the higher the value the lower the commission error of positive samples. The F1 metric is a combination of Recall and Precision metrics which visually evaluates the comprehensive performance of an algorithm. The OA metric is a standard classification accuracy evaluation which visually reflects the detection accuracy of positive and negative samples. The mathematical calculation formulas of the above indicators are as follows:

$$\begin{aligned} Recall &= \frac{TP}{TP+FN} \\ Precision &= \frac{TP}{TP+FP} \\ F1 &= \frac{2 \times Recall \times Precision}{Recall+Precision} \\ OA &= \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \tag{2}$$

where *TP* indicates that the predicted result is a positive sample and the ground truth is also a positive sample. *TN* indicates that the predicted result is a negative sample and the ground truth is also a negative sample. *FP* indicates that the pixel in the ground truth image is a negative sample but is predicted as a positive sample. *FN* indicates that the pixel in the ground truth image is a positive sample but is predicted as a negative sample.

3.5. Implementation Details

We implement our experiment on a computer configured with the Windows-10 operating system, an AMD 5600× CPU, 64 GB DDR4 memory, and a TeslaV100 GPU with 32 GB memory. The deep learning framework used is Pytorch 1.8.1, and the optimizer is the Adam method [46,59]. The initial learning rate is set as 0.001, and the total training epoch is set as 60. After every 10 epochs, the initial learning rate is decreased to 1/10 of the original rate, and the batch size is set to 6. After the model was trained, we used half-precision for model inference with the goal of saving GPU memory and inputting larger sliced images in the model inference stage, thus improving the processing efficiency on large-scale images. For data augmentation methods, we only used random angle rotation, adding Gaussian noise, random scaling, and luminance transformation [56] to improve the generalization ability of the model. The total model training time was 28 h.

3.6. Comparision with other Deep Learning Models

In essence, our deforestation detection method still belongs to the subfield of land cover change detection, and thus it is necessary to compare it with other deep learning

change detection models. We choose some other excellent change detection models that have been published in recent years for comparison. Considering that most of the existing change detection methods use an encode–decoder semantic segmentation architecture, we also select several classical semantic segmentation models for comparison. (The detailed information of compared models can be seen in Table 6).

Table 6. Comparison models.

Method Category	Method Name	Method Description
Change detection models based on deep learning	SiamFCN [36]	SiamFCN uses a fully convolutional neural network for RS image change detection for the first time. The authors completed the pixel-level RS image change detection by image concatenation, feature concatenation, and feature differencing, and verified the effectiveness of a fully convolutional neural network applied to change detection for the first time.
	Unet++ [37]	Unet++ addresses the problem that most existing change detection algorithms use feature layer comparison, which may lead to greater error transfer, and effectively improves accuracy by stacking the bi-temporal images, inputting them into a deep neural network, and then optimizing the output features at different levels and scales.
	STAnet [38]	STAnet proposes a spatial–temporal attention module for building change detection to address the pixel "misregistration" caused by multi-temporal remote sensing image registration. STAnet fused Channel Attention Module and Position Attention Module to obtain state-of-the-art experimental results on the newly proposed building change detection dataset.
	DTCDSCN [39]	DTCDSCN addresses the problem of inaccurate boundaries in existing building extraction models by using two semantic segmentation networks and a siamese change detection network, and then introducing a dual-attention module with focal loss to improve accuracy.
	ESCNet [40]	ESCNet proposed a deep convolutional neural network change detection method based on superpixels for the problem of noise in pixel-level change detection, effectively solving the pixel noise problem and extracting accurate change boundaries.
	SNUNet [41]	SNUNet is a dense feature-linked change detection model for high-resolution remote sensing images. By analyzing the problems of inaccurate boundary segmentation in existing change detection models, it proposes a dense feature-connected deep siamese change detection module, and finally obtains accurate boundaries by designing an ECA module (Ensemble Channel Attention Module) to acquire fine semantic features.
Semantic segmentation models	Unet [60]	Unet was originally designed to solve the medical image segmentation problem, and it used a very elegant encode–decode architecture for the first time to achieve a surprising level of segmentation accuracy by fusing different level semantic features; it provides a very important reference for the subsequent development of semantic segmentation.
	PSPnet [45]	PSPnet mainly proposed a PPM module, which aggregates contextual features at different scales through multiple max-pooling operations, enabling the network to perceive richer semantic features and improve semantic segmentation accuracy.
	DeepLabV3+ [44]	DeepLabV3+ mainly proposes a Spatial Pyramid Pooling Module (ASPP), which obtains contextual features at different scales by the dilation convolution, enabling semantic segmentation models to effectively improve segmentation accuracy with the same computational efficiency.
	HRnet [30]	HRnet designed a high-resolution semantic segmentation network, whose main idea is that the network retains high-resolution feature transfer during the learning process. HRnet achieved state-of-the-art experimental results in the semantic segmentation task and the final segmentation result is excellent.

4. Results

4.1. Visual Evaluation of Deforestation Detection

The deforestation detection results for Hengyang city and Qujing City can be seen in Figures 7 and 8, respectively. It is worth noting that the ground truth polygons are generated from the aforementioned webGIS system. In this section, we will discuss the deforestation detection results extracted by the SiamHRnet-OCR model and the cause of deforestation.

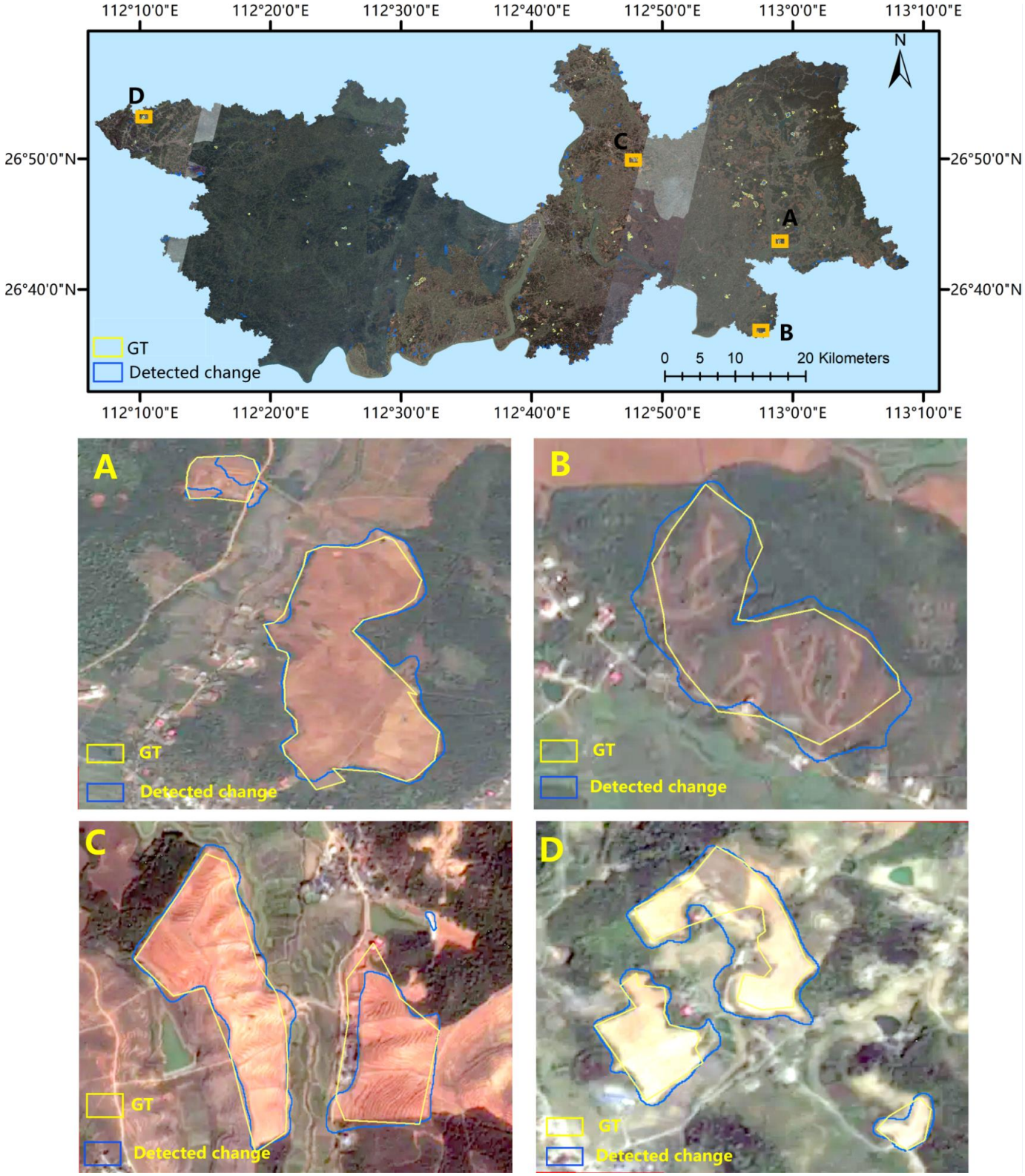


Figure 7. Deforestation detection for Hengyang city.

In Figure 7, on the whole, the area of deforestation detection extracted by SiamHRnet-OCR is 11.24 km^2 (Table 7). Compared with the GT (9.43 km^2), the detected deforestation result of SiamHRnet-OCR was a little overestimated. It is worth noting that the deforestation detection results of SiamHRnet-OCR are not processed by any post-processing

methods. In terms of details, as shown in subregions A, B, C, and D, the boundary of the deforestation areas extracted by the SiamHRnet-OCR is well sketched. Furthermore, to improve the efficiency of visual interpretation, a fast and simplified polygon is used to describe the GT map. Though it seems not to fit very well with the base high-resolution image, using these polygons to evaluate the quantitative accuracy is enough. The shape of the detected change regions extracted by the SiamHRnet-OCR seems better than the GT map, because the training samples proposed in this study are collected from a large region and all of them are carefully examined for quality check; high-quality training samples are very important to the final deforestation detection models. The deforestation detection result for Qujing city is shown in Figure 8:

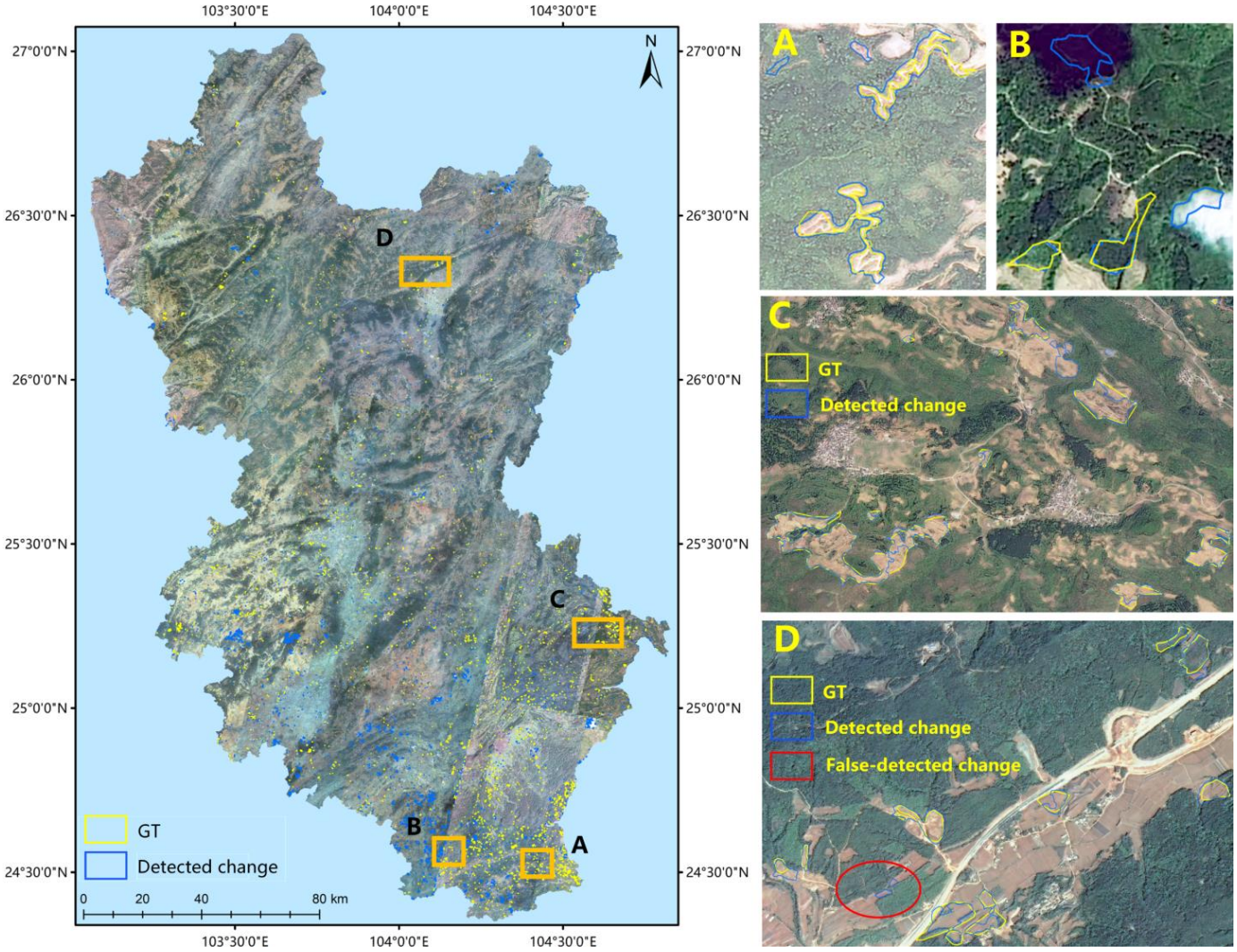


Figure 8. Deforestation detection for Qujing City.

Table 7. Area of deforestation in Hengyang City and Qujing City.

Study Area	GT (km^2)	SiamHRnet-OCR (km^2)	Area Statistics after Removing Commission Alarms (Caused by Cloud and Cloud Shadow)
Hengyang	9.43	11.24	10.05
Qujing	23.46	32.79	25.72

In Figure 8, we can see that deforestation in Qujing City mainly takes place in the southeast region, which is an important area for agricultural cultivation in Qujing City. On the whole, the changing area detected by the SiamHRnet-OCR is 32.79 km^2 (Table 7), while the GT is 23.46 km^2; the difference between them is mainly due to the effect of thick clouds and cloud shadows (as shown in subregion B). In terms of detail, the SiamHRnet-OCR also achieved excellent performance, as shown in subregion A; for example, the newly constructed road boundary is extracted accurately. In subregions C and D, the SiamHRnet-OCR also accurately outlines the boundaries of deforestation regions, but in the red oval region of subregion D two small erroneous polygons are produced, which is because the SiamHRnet-OCR incorrectly identifies water change into bare land as deforestation, possibly causing the area of our deforestation result to be overestimated. The whole deforestation statistical area of each of these two study areas is listed in Table 7.

In Table 7, it can be seen that the area of deforestation detected with the SiamHRnet-OCR is a little larger than the GT of both two cities, which is mainly the effect of clouds and cloud shadows. After removing these commission errors caused by clouds and cloud shadows, the deforestation result of SiamHRnet-OCR is very close to the GT in both two cities. However, there are still a few commission errors in the deforestation results extracted by the SiamHRnet-COR. To further improve the deforestation detection accuracy of SiamHRnet-OCR, an effective post-processing method is using forest products to mask these commission errors. For example, the forest type in GlobaLand30 [49] is a very useful auxiliary method to achieve this. Additionally, though the image acquisition time of these two study areas is different from the training samples, the final deforestation detection result with the SiamHRnet-OCR is still satisfactory. This phenomenon suggests that collecting some samples from the experimental area without needing to consider time consistency is a simple but effective means to improve model performance.

4.2. Quantitative Accuracy Assessment

We calculated Recall, Precision, F1, and OA for accuracy assessment (without removing polygons covered by cloud and cloud shadow). The quantitative accuracy evaluations for Hengyang City and Qujing City are listed in Table 8.

Table 8. Accuracy assessment.

Study Area	Recall (%)	Precision (%)	F1 (%)	OA (%)
Hengyang City	78.97	44.24	56.71	99.85
Qujing City	69.16	44.28	50.52	99.81

We can see that the Recall indicator in the two study areas is high; indicating that most deforestation regions can be captured by the SiamHRnet-OCR in both the Hengyang City and the Qujing City. The precision indicator in the two study areas is relatively low compared with other three indicators, mainly because of the effect of cloud or cloud shadow cover on remote sensing images; these clouds or cloud shadow pixels will be easily identified as changed pixels by SiamHRnet-OCR, increasing the commission errors. The overall accuracy of the two study regions is very high because most pixels in the deforestation detection images are negative pixels, which will greatly affect the OA indicator [12]. Additionally, the F1 indicator in the Qujing City is slightly lower than that of the Hengyang City, mainly because the proportion of cloud pixels in two time-phase images of Qujing City is relatively higher than in Hengyang City. However, the F1 indicators in the two study areas are both larger than 0.5, meaning that the performance of the SiamHRnet-OCR is robust.

4.3. Efficiency Test

Firstly, we slice the bi-temporal large images into small blocks before inputting them into the SiamHRnet-OCR model, because direct input of two bi-temporal large images will

cause some errors, a common example being "out of memory" caused by limited GPU memory. The size of each small block is 4096 × 4096 pixels, which can make full use of the GPU memory resources (the GPU is Tesla V100 with 32 GB memory). Then, deforestation detection is carried out on each small block image. Finally, the predicted deforestation block images are mosaiced to form the final deforestation detection image. The total time consumed in the Hengyang City and theQujing City is shown in Table 9:

Table 9. Accuracy assessment.

Study Area	Image Size (Pixels)	GPU	Time Consumed (s)
Hengyang City	48,448 × 20,202 × 3	TeslaV100	205.1
Qujing City	150,294 × 98,345 × 3	Tesla V100	2805.2

In terms of computational efficiency, we directly used the trained model files without converting the model files into ONNX files or using any other quantization acceleration methods to speed up the deep learning model. The overall computation time can be controlled within 1 h for a very large image with a size of 150,294 × 98,345 × 3 pixels. It can be summarized from this point that with the development of computing hardware devices, and the continuous breakthroughs in the computing power of AI (Artificial Intelligence) chips in the future, the efficiency of model inference will no longer be the bottleneck problem, and thus using high-resolution or very high-resolution remote sensing imagery to monitor deforestation may become the mainstream method in a special region or even on a global scale; we believe that the proposed deforestation detection dataset and the SiamHRnet-OCR model in this study can provide data and method support for the deforestation detection research field.

4.4. Factors in Deforestation

To understand what the main factors in deforestation in the study area are, we choose the Qujing City as an example for analysis. After analyzing each deforestation detection polygon extracted by the SiamHRnet-OCR model (each polygon is analyzed by visual interpretation based on interpretation marks, such as deforestation caused by mining activities, agriculture, or other factors), we find that the types of potential deforestation factors can be divided into five categories: urbanization, infrastructure construction, agriculture, and others. Then we conduct a statistical analysis of the deforestation area of each category, as shown in Table 10.

Table 10. Factors in deforestation.

Type of Deforestation	Proportion (%)
urbanization	5.02
infrastructure construction	25.75
agriculture	45.21
mining activities	20.29
others	3.73

In Table 10, the agriculture factor comprises a dominant proportion of deforestation, possibly due to factors such as growing cash crops. Infrastructure construction also plays an important role in deforestation, reflecting the development strategy of the Chinese government in the southwest region, which is developing infrastructure first to promote economic development. Furthermore, from the perspective of China's recent policy of "common prosperity", strengthening infrastructure development is also a very important way to address the gap between rich and poor regions.

Another interesting question is where deforestation occurs. We first extract the centerlines of all roads (the width of roads is over 4 m) with the D-Linknet model [46]. Then we count the distances from the deforestation regions to the nearest road centerlines, and

finally we make a statistical accounting of all distance values (our deforestation regions are after removing commission alarms caused by clouds and cloud shadows). We found that most of the deforestation occurred in the range of 0–3000 m from roads (Figure 9), which means that human activity plays an important role in deforestation, one of the main reasons being that human activities can be conveniently carried out in these areas.

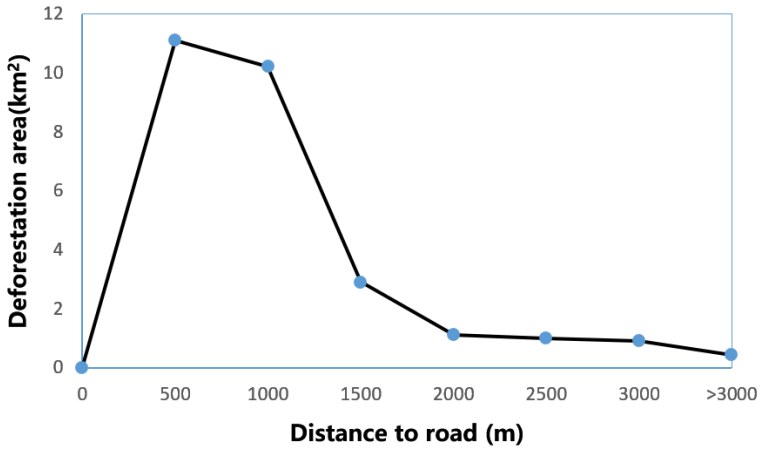


Figure 9. The statistics result for deforestation.

5. Discussion

The deforestation detection results in the Hengyang City and the Qujing City indicate that the boundary of change regions produced by the SiamHRnet-OCR is satisfying, which allows us to address the following questions: What is the feature extraction ability of SiamHRnet-OCR? What are the advantages of the SiamHRnet-OCR versus other deep learning models? What is the advantage of deforestation detection results by the SiamHRnet-OCR over existing deforestation products? To answer these questions, we did some qualitative and quantitative experimental analyses.

5.1. Feature Extraction Ability of the SiamHRnet-OCR

In this study, we proposed a deforestation detection model—SiamHRnet-OCR—to monitor deforestation using high-resolution RS images. To answer the first question (What is the feature extraction ability of SiamHRnet-OCR?), we used a feature visualization methodto help understanding [39]. The deep feature extraction module, the deep feature fusion module, and the OCR refine module in the SiamHRnet-OCR model are visualized at different stages, as shown below.

In Figure 10, we can see how the features change in different layers of SiamHRnet-OCR (the feature map in the above figure is the strongest feature response in the corresponding layer). It clearly indicates that, with the deepening of the model layer, the response of change information in deforestation regions is more and more obvious. It is interesting that, during the whole feature extraction process from Stage 1 to Stage 4, the extracted features are gradually gathered in the change regions and can finally be accurately located in deforestation regions. From the feature fusion layer to the OCR refine layer, the feature response of the "pseudo-change" regions is largely reduced. This phenomenon means high-level semantic features extracted by the OCR refine module have a positive effect on hard-to-classify regions. indicating that the SimaHRnet-OCR model has a strong feature extraction ability to capture the change signal of deforestation, even for ares with subtle changes.

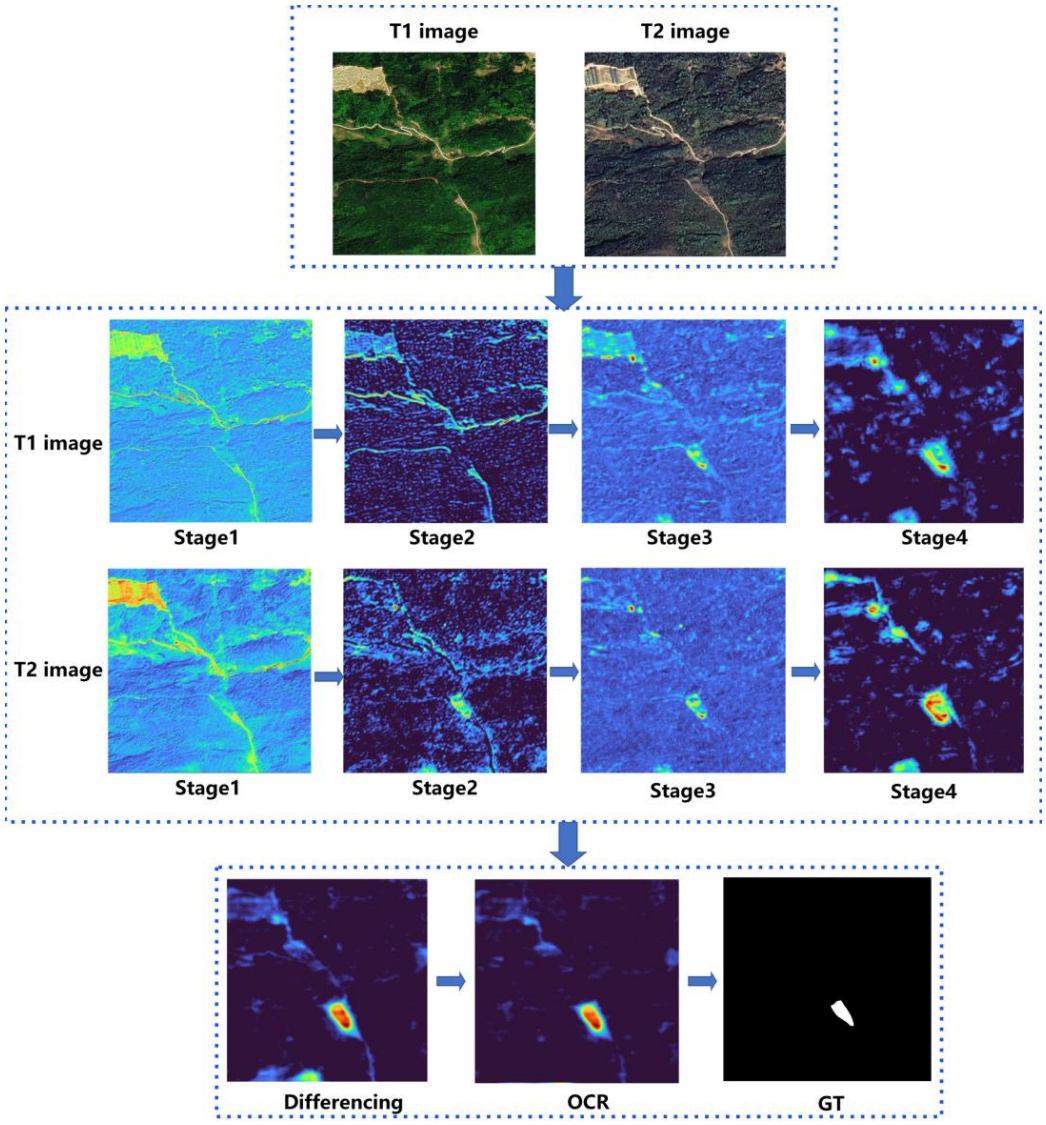


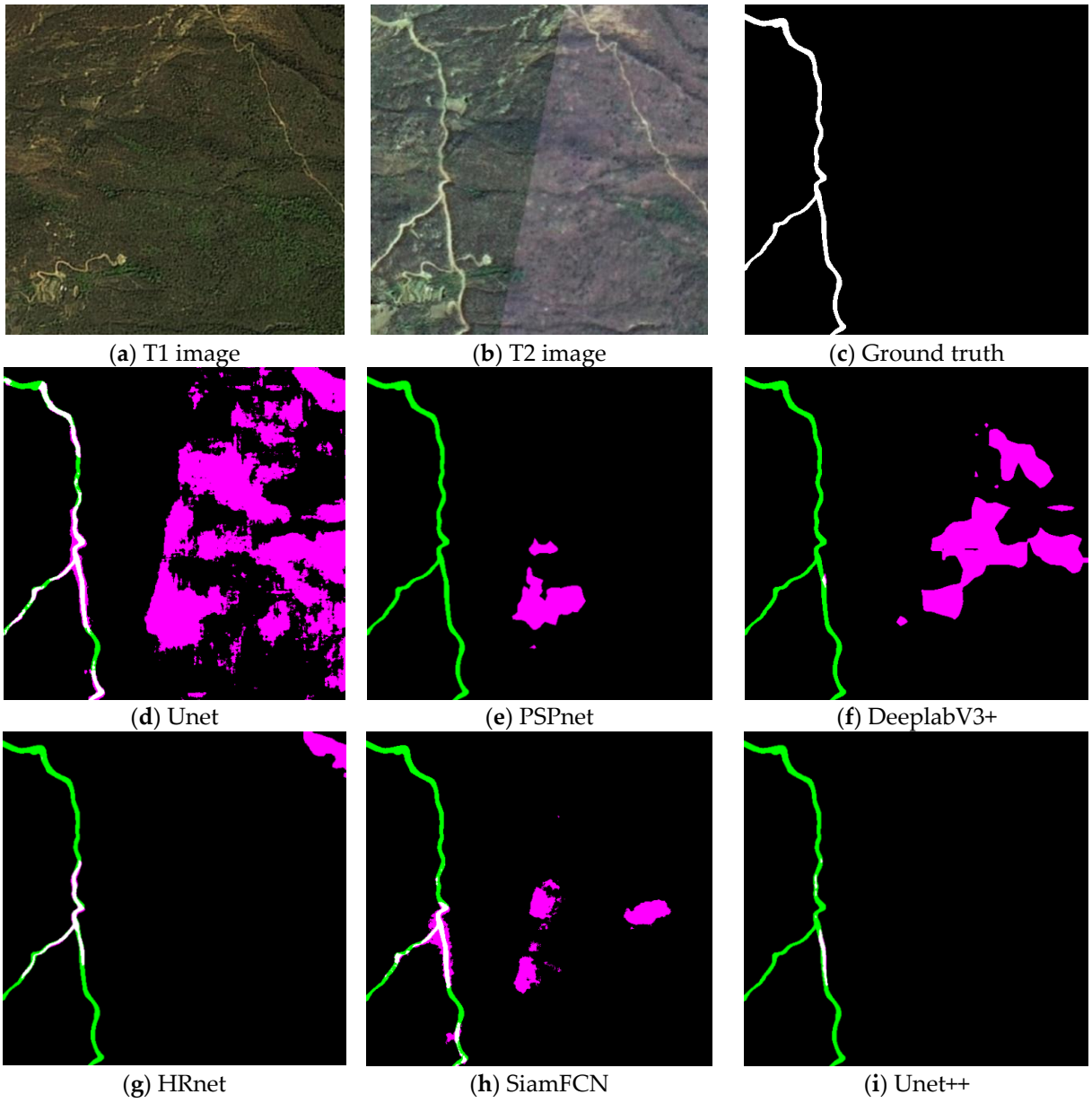
Figure 10. Visualization of a deep feature at a different stages of the SimaHRnet-OCR.

5.2. Comparison with Other Change Detection Methods Based on Deep Learning

To answer the second question (What are the advantages of the SiamHRnet-OCR vs. other deep learning models?), we first discuss the feature extraction ability of different deep learning models for elongated objects. A newly constructed road in the forest was selected for a detailed comparative analysis.

In Figure 11, the deforestation detection result of the semantic segmentation models including Unet, PSPnet, and DeeplabV3+ is relatively worse than those of the change detection models such as Unet++, STAnet, DTCDSCN, ESCNet, SNUNet, etc. From the detailed comparison of deforestation results, there are some commission alarms in Unet, PSPnet, and DeeplablV3+. Essentially, the semantic segmentation models stack two temporal images into a single image with six bands (each time-phase image is three bands), though this means the change detection task can be easily transformed into a semantic segmentation task; the feature extraction ability of the semantic segmentation models may be weaker than the change detection models because the change detection models can explicitly extract the difference between two time-phase images [36]. The depths of SiamFCN and Unet are relatively shallow; thus, their deforestation detection results are relatively worse, because the high-level semantic features extracted by them are not enough to describe the differences in complex scenes. [42] has also demonstrated that deep model depth usually achieves higher accuracy than shallower models. In terms of the spatial resolution

of high-level semantic features, in most of the existing semantic segmentation or change detection models the spatial resolution of high-level semantic features is 1/32 of original input images, such as DeepLabV3+, PSPnet, and SNUnet. Generally speaking, objects in remote sensing images, especially those slender targets such as roads or rivers, high-level semantic features will be lost in deep layers. Therefore, the omission alarms of slender objects in the final detection result will be increased. However, the deforestation detection result produced by the SiamHRnet-OCR indicates that it can accurately capture slender object change, because whether in low-level semantic feature or high-level semantic feature the spatial resolution of semantic features in the SiamHRnet-OCR is always kept as 1/4 of original input images. Such a spatial resolution is suitable for slender object detection; and the above detection result also confirms that such a model structure is effective.



(**a**) T1 image (**b**) T2 image (**c**) Ground truth

(**d**) Unet (**e**) PSPnet (**f**) DeeplabV3+

(**g**) HRnet (**h**) SiamFCN (**i**) Unet++

Figure 11. *Cont.*

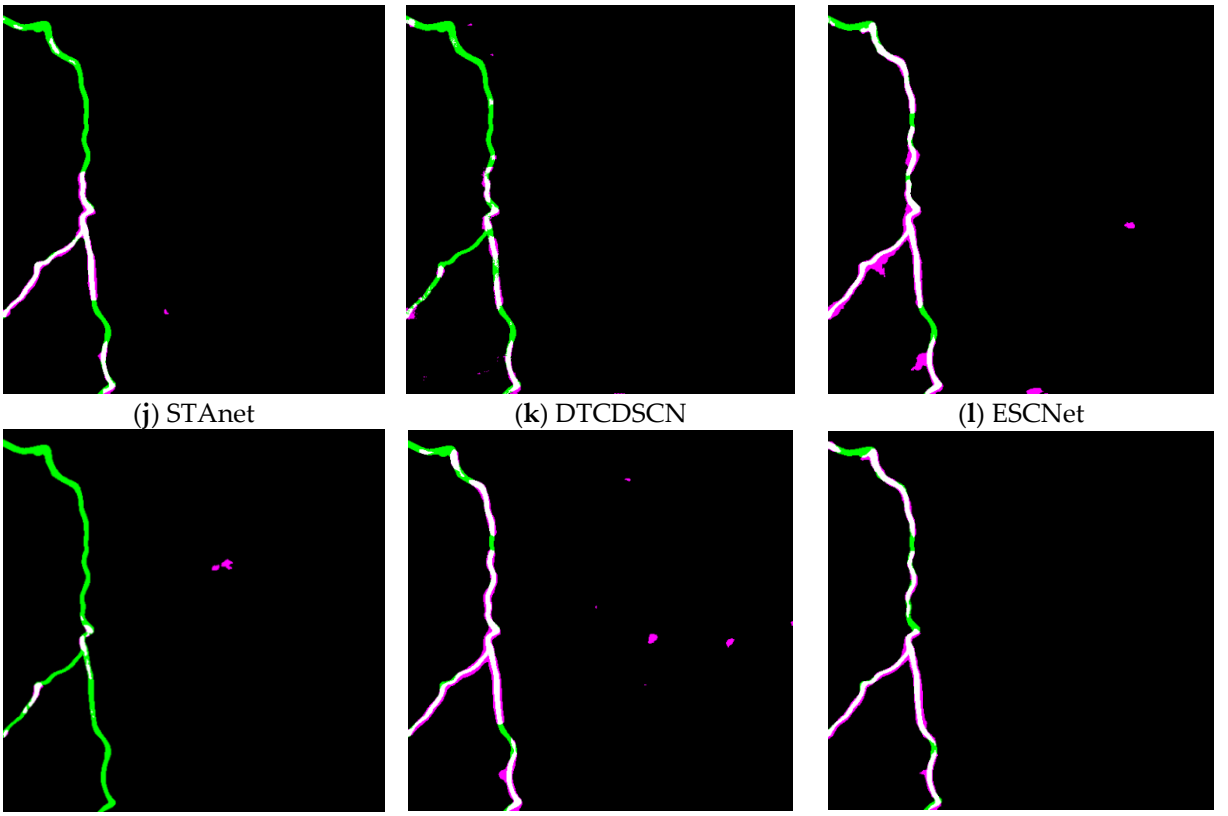
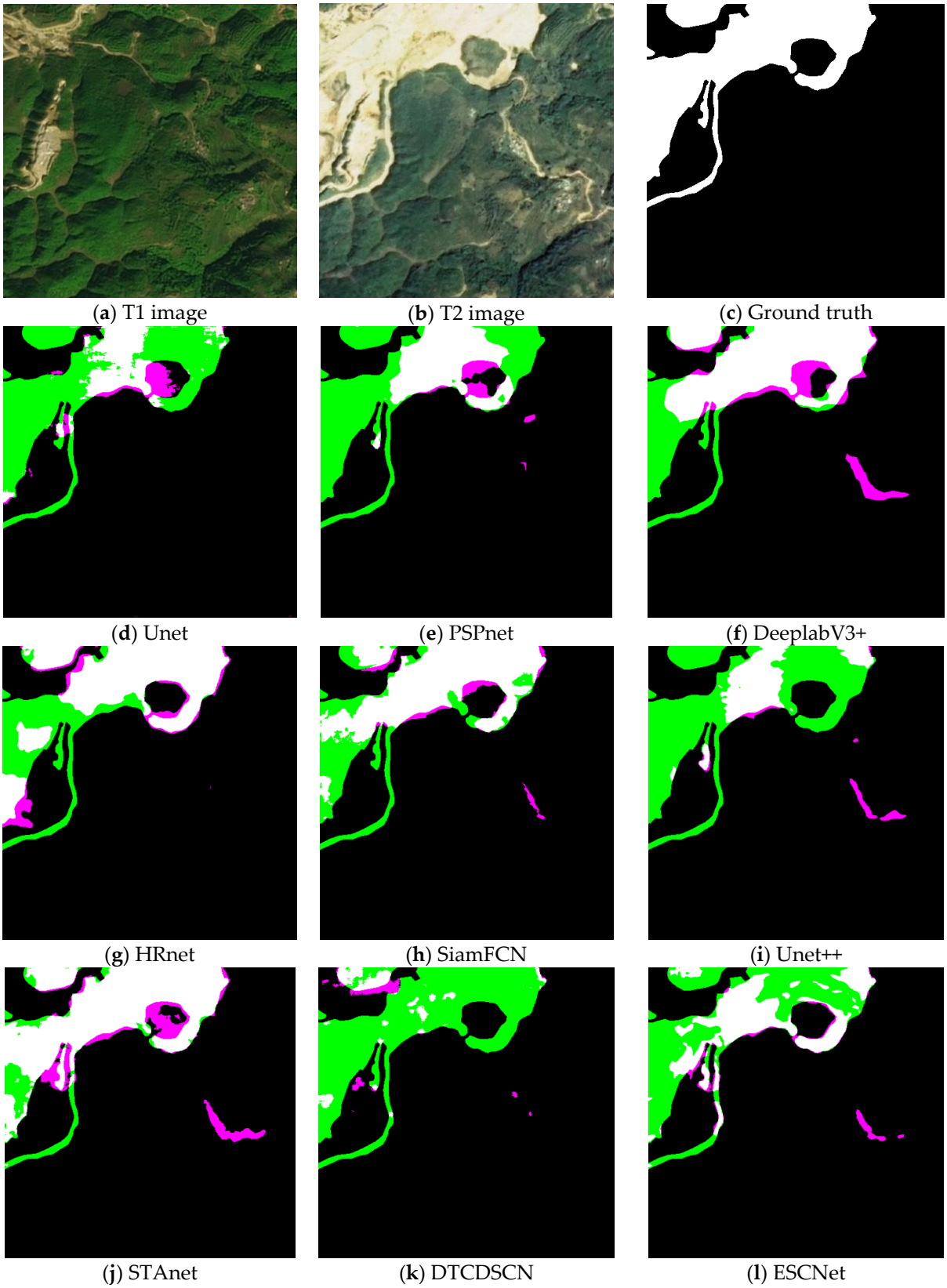


Figure 11. Visual comparison of the change detection results of different methods (the green and magenta color represents missing pixels and false-detected pixels in the change detection result, respectively).

In Figure 11, both SiamHRnet-OCR (concatenation) and SiamHRnet-OCR (differencing) indicate a good visual effect, and the difference between them is negligible. Then how does SiamHRnet-OCR perform in other objects with irregular shapes? An example experiment is shown in Figure 12.

In Figure 12a,b, we can see that the spectral difference between deforestation regions in the bi-temporal images is large, and the shape of the change region is also irregular. As shown in the deforestation detection result of different deep learning models from Figure 12d,o, some omission alarms are produced by the semantic segmentation models, such as in the Unet, PSPnet, and DeeplabV3+ models. It could be that simply stacking the bi-temporal images into a multi-band image will interfere with high-level semantic feature generation [36]. However, this phenomenon also occurred in the change detection models, for example, the Unet++ and DTCDSCN models. This result gives some indication that not all change detection models can achieve excellent performance in monitoring deforestation with high-resolution imagery. In Figure 12j,m, both STAnet and SNUnet achieve relatively good results, but a few omission alarms are still produced in the boundary of change regions of these two methods, especially in the "pseudo-change" regions. As a whole, visually, both SiamHRnet-OCR (concatenation) and SiamHRnet-OCR (differencing) achieved better results than all other models. Additionally, it seems that deforestation detection results based on the SiamHRnet-OCR (differencing) model achieved better visual effects than the SiamHRnet-OCR (concatenation) model, such as in the edge of deforestation regions.



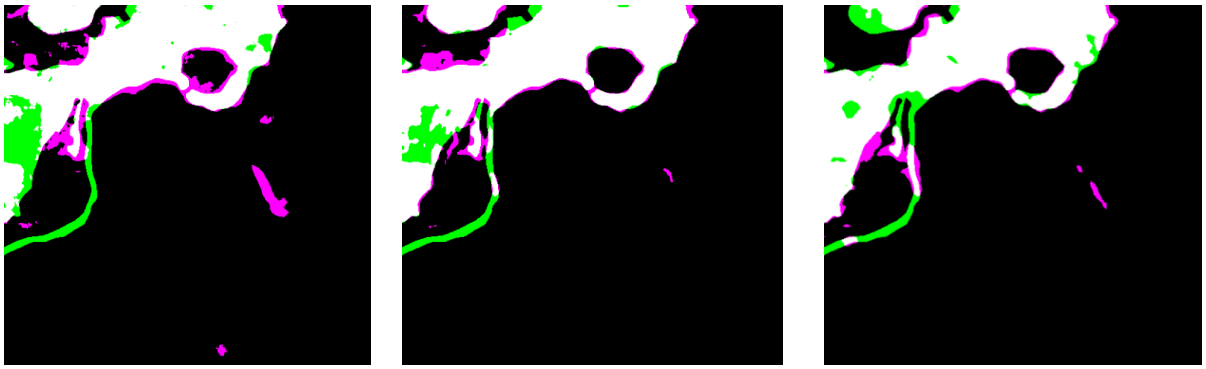
(**a**) T1 image (**b**) T2 image (**c**) Ground truth

(**d**) Unet (**e**) PSPnet (**f**) DeeplabV3+

(**g**) HRnet (**h**) SiamFCN (**i**) Unet++

(**j**) STAnet (**k**) DTCDSCN (**l**) ESCNet

Figure 12. *Cont.*



(**m**) SNUNet (**n**) SiamHRnet-OCR (concatenation) (**o**) SiamHRnet-OCR (differencing)

Figure 12. Visual comparison of the change detection results on irregularly shaped objects (the green and magenta color represents missing pixels and false-detected pixels in the change detection result, respectively).

We have qualitatively compared and analyzed the change detection results of different models; we have also quantitatively evaluated all the models using the quantitative accuracy evaluation metrics mentioned in Section 3.4. The detection accuracies of different models are shown in Table 11.

Table 11. Model comparison (The FLOPs, Parameter, and Inference time are calculated on two images (each image is 512 × 512 × 3 pixels), the hardware is RTX3090 with 24 GB memory, CUDA11.1).

Model	Recall (%)	Precision (%)	F1 (%)	OA (%)	FLOPs	Parameter	Inference Time
Unet	73.51	30.45	37.95	96.56	8.81 G	0.90 M	0.085 s
PSPnet	83.84	49.35	57.67	9855	541.64 G	65.70 M	0.247 s
DeepLabV3+	90.45	44.06	55.26	98.22	332.11 G	58.04 M	0.238 s
HRnet	83.78	53.35	59.59	98.45	207.32 G	40.20 M	0.188 s
SiamFCN	85.05	46.65	54.50	97.73	37.81 G	1.35 M	0.103 s
Unet++	82.07	53.50	59.00	98.40	279.69 G	9.16 M	0.128 s
STAnet	88.58	54.60	63.73	98.70	102.87 G	16.89 M	0.103 s
DTCDSCN	68.85	55.10	56.18	98.17	163.57 G	41.07 M	0.142 s
ESCNet	83.84	48.19	55.35	98.14	353.21 G	5.13 M	0.367 s
SNUNet	81.62	62.81	65.92	98.71	438.66 G	12.03 M	0.197 s
SiamHRnet-OCR (concatenation)	85.00	60.25	66.13	98.95	628.66 G	50.51 M	0.252 s
SiamHRnet-OCR (differencing)	84.05	**64.82**	**68.92**	**98.98**	306.62 G	40.20 M	0.236 s

In Table 11, among all deep learning models, the Precision, F1, and OA indicators of the SiamHRnet-OCR (differencing) model achieve the highest scores. The F1 indicator of the SiamHRnet-OCR (concatenation) model was slightly lower than SiamHRnet-OCR (differencing). Moreover, the complexity comparison between SiamHRnet-OCR (differencing) and SiamHRnet-OCR (concatenation) indicates that SiamHRnet-OCR (differencing) has fewer parameters and a faster inference speed than SiamHRnet-OCR (concatenation), for example, the FLOPs indicator of SiamHRnet-OCR (differencing) is only 48.77% of SiamHRnet-OCR (concatenation), and the model Parameters of SiamHRnet-OCR (differencing) is 79.58% of SiamHRnet-OCR (concatenation).

Although compared with the lightweight models, such as Unet, SiamFCN, and SNUNet, the inference time of SiamHRnet-OCR (differencing) is slower, the Precision, F1, and OA indicators show that SiamHRnet-OCR (differencing) has higher accuracy results than them—for instance, the F1 indicator of SiamHRnet-OCR (differencing) is 3.0% higher than the SNUnet model. Moreover, compared with other relatively heavyweight models, such as PSPnet, DeepLabV3+, and ESCNet, the SiamHRnet-OCR (differencing)

model has a faster inference speed. In addition, our experiment also confirms the finding of recent research [61] that it is important to keep high-resolution features forwarding in the training process to acquire rich and useful context semantic information, which can improve the model detection accuracy of slender objects or other complex objects.

5.3. Comparison with an Existing Forest Change Product

What is the advantage of deforestation detection results by SiamHRnet-OCR over other existing deforestation products? To answer this question, we selected the Hengyang City for comparison. The current highest resolution forest change detection product covering large regions available was proposed by [4], namely GFC-V1.8 (Hansen Global Forest Change V1.8), which has a 30 m spatial resolution. To maintain time consistency between GFC-V1.8 and our result, we selected the 2019 global forest loss product of GFC-V1.8 for comparison. The GFC-V1.8 and the SiamHRnet deforestation detection results are shown in Figure 13.

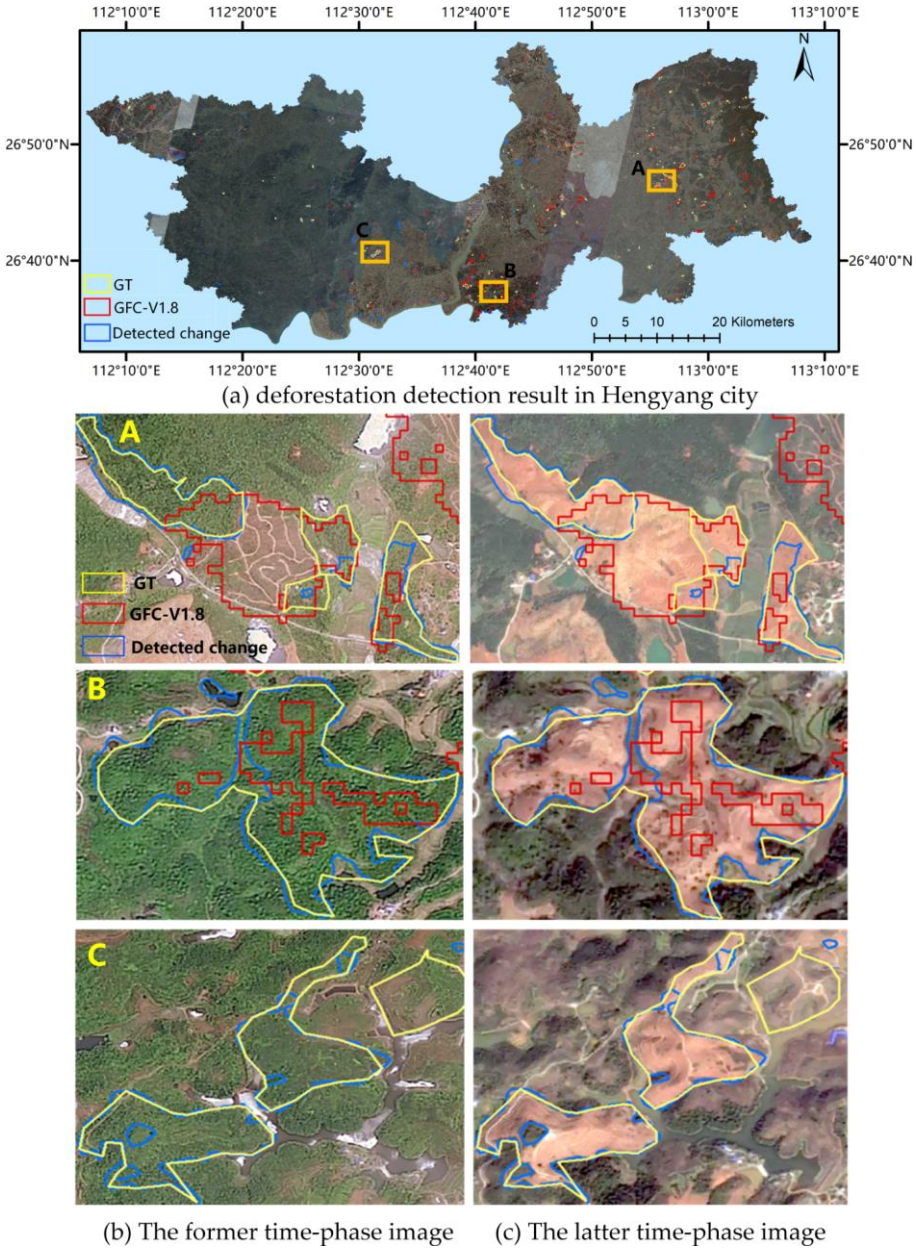


Figure 13. Forest change detection results.

As shown in the sub-regions A and B in Figure 13, the deforestation boundaries detected by the SiamHRnet-OCR are accurate and almost identical to the GT boundaries. Though the spatial resolution of GFC-V1.8 is 30 m, it can still accurately locate forest change. However, GFC-V1.8 produces a few omission alarms in the central region of subregion A, because this region was covered by weeds in the former time-phase image with high NDVI values, causing it to be incorrectly considered as forest cover. By contrast, SiamHRnet-OCR can effectively distinguish between grass and forest on 2 m high-resolution remote sensing images. In sub-region C, the GFC-V1.8 product did not detect deforestation, perhaps due to cloud cover or missing image data.

The quantitative accuracy comparison between the SiamHRnet-OCR and the GFC-V1.8 can be seen in Table 12. It indicates that all four accuracy assessment indicators of the deforestation detection result from SiamHRnet-OCR are higher than those of the GFC-V1.8 product; in particular, the F1 indicator of SiamHRnet-OCR is 40.75% higher than GFC-V1.8. In terms of spatial detail, the visual effect of our results is also relatively superior. It is worth mentioning that the GFC-V1.8 product is produced on Landsat imagery with 30 m spatial resolution, and thus the comparison between the SiamHRnet-OCR and the GFC-V1.8 is not entirely fair so it is not possible to say that the deforestation detection result based on SiamHRnet-OCR is better than GFC-V1.8 product. However, the above comparison further confirms that deep learning methods are a good choice to achieve high-precision deforestation detection with high-resolution remote sensing imagery.

Table 12. Accuracy assessment.

Methods	Recall (%)	Precision (%)	F1 (%)	OA (%)
GFC-V1.8	18.15	14.24	15.96	99.76
SiamHRnet-OCR	78.97	44.24	56.71	99.85

Statistical analysis results indicate that the forest loss area detected by GFC-V1.8 in the Hengyang City is 6.05 km^2, which is significantly lower than the GT (9.43 km^2), while the total deforestation area detected by SiamHRnet-OCR is 11.24 km^2, which is slightly larger than the GT. There are three possible reasons to explain this difference: (1) GFC-V1.8 was produced on 30 m Landsat imagery, which is a relatively coarse spatial resolution, and may be not suitable for high-precision deforestation detection; (2) some deforestation regions with slight spectral change cannot be captured by GFC-V1.8; and (3) a few commission errors were produced by the SiamHRnet-OCR, e.g. water changing into bare land was regarded as deforestation.

5.4. Limitations

With the help of a large quantity of high-quality deforestation training samples, deforestation detection with high-resolution imagery has been investigated in this study and proves the feasibility and efficiency of the SiamHRnet-OCR model in deforestation detection tasks. However, there is still room for further improvement.

(1) In this newly proposed deforestation detection dataset, the SimaHRnet-OCR achieved excellent performance; however, further experiments are needed to verify whether SimaHRnet-OCR is still the best model on other change detection training datasets.

(2) The SiamHRnet-OCR model can only be applied to two bi-temporal image change detections so far, and the next step is to extend the trained deep learning model for long time-series deforestation detection tasks.

(3) The SiamHRnet-OCR model produced a few omission errors in cloud and cloud shadowing covered regions. To improve the model accuracy, the automatic or semi-automatic cloud and cloud shadow masking algorithms can be used as the pre-processing means to further improve detection accuracy [62].

6. Conclusions

In this paper, a new deforestation detection dataset was proposed, with a spatial resolution of 2 m. This dataset was generated over a large region, namely the whole region of the Yangtze River Economic Zone of China. In addition, we designed a new deforestation detection model based on a deep learning method: SiamHRnet-OCR. Compared with other deep learning models, SiamHRnet-OCR achieved excellent performance with the new deforestation detection dataset. In the future, we will also test the potential ability of the SiamHRnet-OCR on other landcover datasets and continue to increase the number of deforestation datasets over a larger region.

Author Contributions: Conceptualization, Z.W. (Zhipan Wang) and Q.Z.; Methodology, W.P. and D.L.; Validation and writing the original draft, Z.W. (Zhipan Wang) and Q.Z.; Formal analysis, Q.Z., Z.W. (Zhongwu Wang) and X.L.; Writing, review, editing, and supervision Z.W. (Zhipan Wang), W.P., D.L. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shenzhen Science and Technology Innovation Project (No. ZDSYS20210623091808026), and supported in part by the National Natural Science Foundation of China (General Program, No. 42071351), the National Key Research and Development Program of China (No. 2020YFA0608501), the Chongqing Science and Technology Bureau technology innovation and application development special (cstc2021jscx-gksb0116).

Acknowledgments: We would like to thank the Land Satellite Remote Sensing Application Center of China for providing high-resolution images.

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

References

1. Nesha, M.K.; Herold, M.; De, S.V.; Duchelle, A.E.; Christopher, M.; Anne, B.; Monica, G. An assessment of data sources, data quality and changes in national forest monitoring capacities in the Global Forest Resources Assessment 2005–2020. *Environ. Res. Lett.* **2021**, *16*, 54029. [CrossRef]
2. Achard, F.; Eva, H.D.; Stibig, H.J.; Mayaux, P.; Gallego, J.; Richards, T.; Malingreau, J.P. Determination of deforestation rates of the world's humid tropical forests. *Science* **2002**, *297*, 999–1002. [CrossRef] [PubMed]
3. Curtis, P.G.; Slay, C.M.; Harris, N.L.; Tyukavina, A.; Hansen, M.C. Classifying drivers of global forest loss. *Science* **2018**, *361*, 1108–1111. [CrossRef]
4. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853. [CrossRef] [PubMed]
5. Tang, L.; Wang, H.H.; Li, L.; Yang, K.T.; Mi, Z.F. Quantitative models in emission trading system research: A literature review. *Renew. Sustain. Energy Rev.* **2020**, *132*, 110052. [CrossRef]
6. Moutinho, P.; Guerra, R.; Azevedo-Ramos, C.; Kapuscinski, A.R.; Frumhoff, P.C. Achieving zero deforestation in the Brazilian Amazon: What is missing? *Elem. Sci. Anthr.* **2016**, *4*, 456–476. [CrossRef]
7. Andrade, F.W.C.; Pinto, T.I.; Moreira, L.d.S.; da Ponte, M.J.M.; Lobato, T.d.C.; de Sousa, J.T.R.; Moutinho, V.H.P. The Legal Roundwood Market in the Amazon and Its Impact on Deforestation in the Region between 2009–2015. *Forests* **2022**, *13*, 558. [CrossRef]
8. Lima, M.; Santana, D.C.; Junior, I.C.M.; Costa, P.M.C.d.; Oliveira, P.P.G.d.; Azevedo, R.P.d.; Silva, R.d.S.; Marinho, U.d.F.; Silva, V.d.; Souza, J.A.A.d.; et al. The "New Transamazonian Highway": BR-319 and Its Current Environmental Degradation. *Sustainability* **2022**, *14*, 823. [CrossRef]
9. Alzu'bi, A.; Alsmadi, L. Monitoring deforestation in Jordan using deep semantic segmentation with satellite imagery. *Ecol. Inform.* **2022**, *70*, 101745. [CrossRef]
10. Cardille, J.A.; Perez, E.; Crowley, M.A.; Wulder, M.A.; White, J.C.; Hermosilla, T. Multi-sensor change detection for within-year capture and labelling of forest disturbance. *Remote Sens. Environ.* **2022**, *268*, 112741. [CrossRef]
11. Chen, S.; Woodcock, C.E.; Bullock, E.L.; Arévalo, P.; Torchinava, P.; Peng, S.; Olofsson, P. Monitoring temperate forest degradation on Google Earth Engine using Landsat time series analysis. *Remote Sens. Environ.* **2021**, *265*, 112648–112666. [CrossRef]
12. De Bem, P.; de Carvalho Junior, O.; Fontes Guimarães, R.; Trancoso Gomes, R. Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 901. [CrossRef]
13. Feng, Z.; Rui, S.; Liheng, Z.; Ran, M.; Chengquan, H.; Xiaoxi, Z.; Mengyu, W. Monthly mapping of forest harvesting using dense time series Sentinel-1 SAR imagery and deep learning. *Remote Sens. Environ.* **2022**, *269*, 112822. [CrossRef]
14. Irvin, J.; Sheng, H.; Ramachandran, N.; Sonja, J.-Y.; Sharon, Z.; Kyle, S.; Rose, R. Forestnet: Classifying drivers of deforestation in indonesia using deep learning on satellite imagery. *arXiv* **2020**, arXiv:2011.05479. [CrossRef]

15. John, D.; Zhang, C. An attention-based U-Net for detecting deforestation within satellite sensor imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102685. [CrossRef]
16. Woodcock, C.E.; Allen, R.; Anderson, M.; Belward, A.; Bindschadler, R.; Cohen, W.; Gao, F.; Goward, S.N.; Helder, D.; Helmer, E.; et al. Free access to Landsat imagery. *Science* **2008**, *320*, 1011. [CrossRef] [PubMed]
17. Fortin, J.A.; Cardille, J.A.; Perez, E. Multi-sensor detection of forest-cover change across 45 years in Mato Grosso, Brazil. *Remote Sens. Environ.* **2020**, *238*, 210–266. [CrossRef]
18. Negassa, M.D.; Mallie, D.T.; Gemeda, D.O. Forest cover change detection using Geographic Information Systems and remote sensing techniques: A spatio-temporal study on Komto Protected forest priority area, East Wollega Zone, Ethiopia. *Environ. Syst. Res.* **2020**, *9*, 1. [CrossRef]
19. Oduro Appiah, J.; Agyemang-Duah, W.; Sobeng, A.K.; Kpienbaareh, D. Analysing patterns of forest cover change and related land uses in the Tano-Offin forest reserve in Ghana: Implications for forest policy and land management. *Trees For. People* **2021**, *5*, 100105–100120. [CrossRef]
20. Rahimi-Ajdadi, F.; Khani, M. Remote sensing-based detection of tea land losses: The case of Lahijan, Iran. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100568–100583. [CrossRef]
21. Tan, B.; Masek, J.G.; Wolfe, R.; Gao, F.; Huang, C.; Vermote, E.F.; Sexton, J.O.; Ederer, G. Improved forest change detection with terrain illumination corrected Landsat images. *Remote Sens. Environ.* **2013**, *136*, 469–483. [CrossRef]
22. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]
23. Ye, S.; Rogan, J.; Zhu, Z.; Eastman, J.R. A near-real-time approach for monitoring forest disturbance using Landsat time series: Stochastic continuous change detection. *Remote Sens. Environ.* **2021**, *252*, 112167. [CrossRef]
24. Chen, H.; Zeng, Z.; Wu, J.; Peng, L.; Lakshmi, V.; Yang, H.; Liu, J. Large Uncertainty on Forest Area Change in the Early 21st Century among Widely Used Global Land Cover Datasets. *Remote Sens.* **2020**, *12*, 3502. [CrossRef]
25. Maretto, R.V.; Fonseca, L.M.G.; Jacobs, N.; Körting, T.S.; Bendini, H.N.; Parente, L.L. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 771–775. [CrossRef]
26. Bragagnolo, L.d.; Silva, R.V.; Grzybowski, J.M.V. Amazon forest cover change mapping based on semantic segmentation by U-Nets. *Ecol. Inform.* **2021**, *62*, 101279. [CrossRef]
27. Rakshit, S.; Debnath, S.; Mondal, D. Identifying land patterns from satellite imagery in amazon rainforest using deep learning. *arXiv* **2018**, arXiv:1809.00340. [CrossRef]
28. Lee, S.H.; Han, K.J.; Lee, K.; Kwang-Jae, L. Classification of landscape affected by deforestation using high-resolution remote sensing data and deep-learning techniques. *Remote Sens.* **2020**, *12*, 3372. [CrossRef]
29. De Andrade, R.B.; Mota, G.L.A.; da Costa, G.A.O.P. Deforestation Detection in the Amazon Using DeepLabv3+ Semantic Segmentation Model Variants. *Remote Sens.* **2022**, *14*, 4694. [CrossRef]
30. Tao, A.; Sapra, K.; Catanzaro, B. Hierarchical Multi-Scale Attention for Semantic Segmentation. *arXiv* **2020**, arXiv:2005.10821. [CrossRef]
31. Seo, D.K.; Kim, Y.H.; Eo, Y.D.; Lee, M.H.; Park, W.Y. Fusion of SAR and Multispectral Images Using Random Forest Regression for Change Detection. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 401. [CrossRef]
32. Ortega, A.M.; Queiroz, F.R.; Nigri, H.P.; Claudio, A.D.A. Evaluation of deep learning techniques for deforestation detection in the Brazilian Amazon and cerrado biomes from remote sensing imagery. *Remote Sens.* **2020**, *12*, 910. [CrossRef]
33. Zhang, B.; Mu, H.; Gao, M.; Haiming, N. A Novel Multi-Scale Attention PFE-UNet for Forest Image Segmentation. *Forests* **2021**, *12*, 937. [CrossRef]
34. Taquary, E.C.; Fonseca, L.G.M.; Maretto, R.V.; Hugo, N.B. Detecting Clearcut Deforestation Employing Deep Learning Methods and SAR Time Series. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4520–4523. [CrossRef]
35. Tovar, P.; Adarme, M.O.; Feitosa, R.Q. Deforestation detection in the amazon rainforest with spatial and channel attention mechanisms. *Remote Sens. Spat. Inf. Sci.* **2021**, *XLIII-B3-2021*, 851–885. [CrossRef]
36. Daudt, R.C.; Saux, B.L.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067. [CrossRef]
37. Peng, D.F.; Zhang, Y.J.; Guan, H.Y. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet plus. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]
38. Chen, H.; Shi, Z. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]
39. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 811–815. [CrossRef]
40. Zhang, H.; Lin, M.; Yang, G.; Zhang, L. ESCNet: An End-to-End Superpixel-Enhanced Change Detection Network for Very-High-Resolution Remote Sensing Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 28–42. [CrossRef] [PubMed]
41. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007805. [CrossRef]
42. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

43. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.X.; Wang, W.J.; Zhu, Y.K.; Pang, R.M.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/Cvf International Conference on Computer Vision (Iccv 2019), Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324. [CrossRef]
44. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018, 15th European Conference, Munich, Germany, 8–14 September 2018; pp. 833–851. [CrossRef]
45. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239. [CrossRef]
46. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–1924. [CrossRef]
47. Yuan, Y.; Chen, X.; Wang, J. Object-Contextual Representations for Semantic Segmentation. In Proceedings of the Computer Vision—ECCV 2020, 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 173–190. [CrossRef]
48. Zhang, Y.J.; Zhang, L.J.; Wang, H.; Wang, Y.Y.; Ding, J.Q.; Shen, J.S.; Wang, Z.; Liu, Y.L.; Liang, C.Y.; Li, S.C. Reconstructing deforestation patterns in China from 2000 to 2019. *Ecol. Model.* **2022**, *465*, 465–477. [CrossRef]
49. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]
50. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
51. Zhang, M.; Shi, W. A Feature Difference Convolutional Neural Network-Based Change Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [CrossRef]
52. Ji, S.P.; Shen, Y.Y.; Lu, M.; Zhang, Y.J. Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
53. Shen, L.; Lu, Y.; Chen, H.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Lv, S.; Jiang, B. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sens.* **2021**, *13*, 5094. [CrossRef]
54. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]
55. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]
56. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.S.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890. [CrossRef]
57. Wen, D.; Huang, X.; Bovolo, F.; Li, J.; Ke, X.; Zhang, A.; Benediktsson, J.A. Change Detection From Very-High-Spatial-Resolution Optical Remote Sensing Images: Methods, applications, and future directions. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 68–101. [CrossRef]
58. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. *arXiv* **2022**, arXiv:2209.08575.
59. Zhu, Y.; Newsam, S. DenseNet for dense flow. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 790–794. [CrossRef]
60. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
61. Yuan, Y.; Rao, F.; Lang, H.; Lin, W.; Zhang, C.; Chen, X.; Wang, J. HRFormer: High-Resolution Transformer for Dense Prediction. *arXiv* **2021**, arXiv:2110.09408.
62. Li, Z.; Shen, H.; Cheng, Q.; Liu, Y.; You, S.; He, Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 197–212. [CrossRef]

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.