




Article

The Effect of Negative Samples on the Accuracy of Water Body Extraction Using Deep Learning Networks

Jia Song^{1,2,*}  and Xiangbing Yan^{1,3}

¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

² Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

³ School of Resource and Environmental Science, Wuhan University, Wuhan 430072, China

* Correspondence: songj@igsrr.ac.cn

Abstract: Water resources are important strategic resources related to human survival and development. Water body extraction from remote sensing images is a very important research topic for the monitoring of global and regional surface water changes. Deep learning networks are one of the most effective approaches and training data is indispensable for ensuring the network accurately extracts water bodies. The training data for water body extraction includes water body samples and non-water negative samples. Cloud shadows are essential negative samples due to the high similarity between water bodies and cloud shadows, but few studies quantitatively evaluate the impact of cloud shadow samples on the accuracy of water body extraction. Therefore, the training datasets with different proportions of cloud shadows were produced, and each of them includes two types of cloud shadow samples: the manually-labeled cloud shadows and unlabeled cloud shadows. The training datasets are applied on a novel transformer-based water body extraction network to investigate how the negative samples affect the accuracy of the water body extraction network. The evaluation results of Overall Accuracy (OA) of 0.9973, mean Intersection over Union (mIoU) of 0.9753, and Kappa of 0.9747 were obtained, and it was found that when the training dataset contains a certain proportion of cloud shadows, the trained network can handle the misclassification of cloud shadows well and more accurately extract water bodies.



Citation: Song, J.; Yan, X. The Effect of Negative Samples on the Accuracy of Water Body Extraction Using Deep Learning Networks. *Remote Sens.* **2023**, *15*, 514. <https://doi.org/10.3390/rs15020514>

Academic Editor: Konstantinos Topouzelis

Received: 8 December 2022

Revised: 8 January 2023

Accepted: 13 January 2023

Published: 15 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/>).

Keywords: water body extraction; deep learning; negative sample; cloud shadow interference; semantic segmentation

1. Introduction

Water resources play an important role in nature, providing necessary resources for human life, industry production and agriculture planting. Since remote sensing satellites have the advantages of large-scale detection of land surface, many remote sensing images have been applied in water body extraction and monitoring the change in water resources [1,2]. Based on the remote sensing images, various methods [3–6] are developed to extract water bodies from remote sensing images, and deep learning is the most popular approach among them [7–10].

The deep learning approach shows great potential in large-scale automatic image classification tasks due to its strong generalization ability. The main deep learning networks used in water body extraction are semantic segmentation networks. FCN [11–14], U-Net [15–19], DeepLab [20–22] and SegNet [23–25] are commonly used semantic segmentation networks to extract water bodies. In addition, some studies focus on the improvement based on them and achieve better accuracy of water body extraction [26–30]. While training samples also play an important role on accuracy improvement, since the deep learning is a data-driven approach and the accuracy heavily relies on the training samples. The existing efforts on training samples include replacing the low-resolution images with the higher-resolution

images to achieve fine water body extraction [31–33] and the utilization of data augmentation techniques [34–36] to compensate for insufficient samples. The main attention of these studies is paid to positive samples for improving the precision of water bodies.

However, the training samples should include not only positive samples but also negative samples. Negative samples in water body extraction refers to the samples that contains non-water objects. With the negative samples, the deep learning networks may better learn the discriminative features of water bodies. This cognition is formed based on the study [37], which proposes a concept of deep feature space for classification tasks. Based on the deep feature space, a gap between the per-class distributions of the training and testing vectors is accurately described. The bigger gap means higher separability of different classes, which helps classifiers to better distinguish different classes. The negative samples may be beneficial for forming the high separability of per-class distributions, thereby obtaining high accuracy.

In water body extraction, cloud shadows can be indispensable negative samples since cloud shadows are easily misclassified as water bodies due to the similarity between them. The existing efforts have still not well addressed this issue. Instead, they usually use cloud-free images or utilize extra cloud-shadow masks in water body extraction [38,39]. The cloud-free images are obtained by directly selecting non-cloud images or by using image composition algorithms. This approach is obviously limited by the situation of the cloud coverage and is not suitable for the areas that are often covered by clouds. When the networks are trained with cloud-free samples, they cannot well distinguish the water bodies and cloud shadows, which leads to the misclassification of cloud shadows. The approach to utilize cloud-shadow masks for eliminating cloud shadows is less efficient since it requires two stages of mask generation and water body extraction. Moreover, the accuracy of cloud shadow masks obtained by Fmask [40] or Tmask [41] algorithms is not very high; thus, they may not effectively avoid the misclassification of cloud shadows.

Therefore, considering that the misclassification of cloud shadows is not well addressed and negative samples in deep learning are not paid significant attention, this study investigates the impact of cloud shadow negative samples on the accuracy of water body extraction, especially for the low accuracy raised by the cloud shadows being misclassified as waters. Moreover, the impact of negative samples on the precision of water boundaries and the water bodies being missed are also involved. We thus designed two groups of training data. Both these groups are based on the same set of Sentinel-2 images, but the difference of the two groups is the labeled aspects. In one group, only water bodies are labeled, and in the other group, both water bodies and cloud shadows are labeled. Each group of the training data further consists of two training datasets with different proportions of cloud shadows, and these training datasets, respectively, are used to train a novel vision transformer-based deep learning network for extracting water bodies. By analyzing the extraction results, we can find how the cloud-shadow negative samples affect the accuracy of water body extraction.

2. Materials and Methods

2.1. Study Area

The western Tibetan Plateau is the study area in this work (Figure 1). It is located at 82.0° E to 92.3° E and 30.3° N to 36.3° N. There are approximately 1500 lakes with an area greater than 1 km² in this region [42]. The region affects the global environment and is an important scientific research area in the fields of hydrology, climate and geography [43]. Thus, monitoring the dynamic changes in water resources in this region has very important research value [44,45]. However, due to the particularity of climate and terrain, this region is often covered by clouds, and the clear-sky images are difficult to obtain in this region. Therefore, the study area is suitable to explore how cloud-shadow negative samples affect the accuracy of water body extraction, and is a challenging area to accurately extract water bodies.

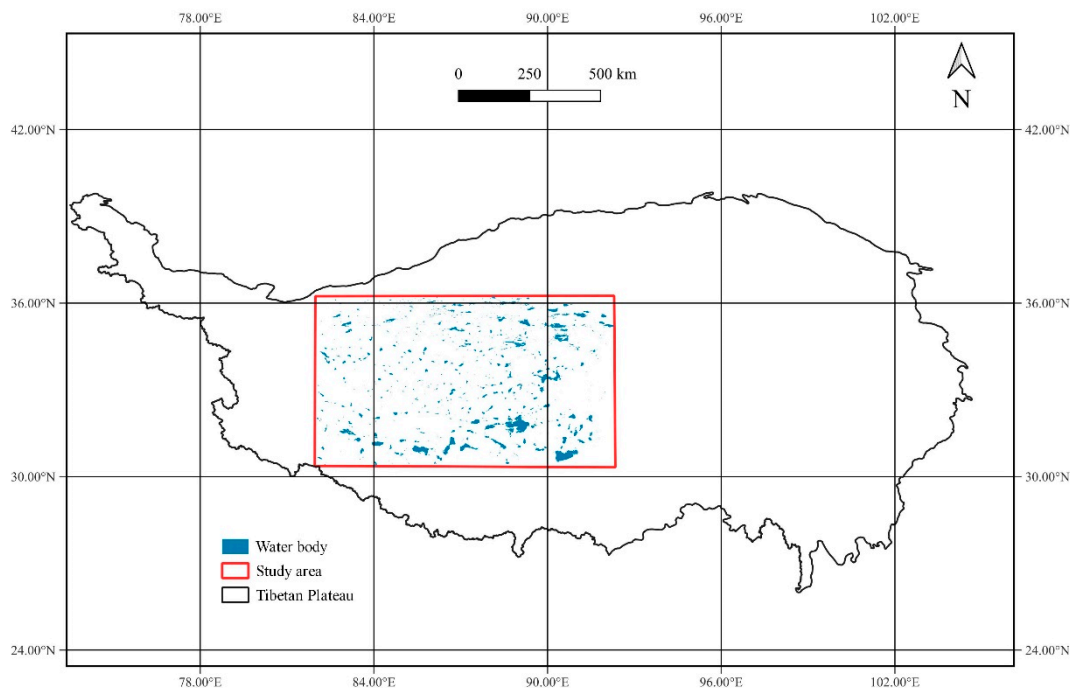


Figure 1. The location of the study area.

2.2. Data Sources

Sentinel-2 imagery is selected as the main data source in this study due to its high spatio-temporal resolution and image quality [46]. The Sentinel-2 image contains 13 spatial resolution bands of 10 m, 20 m and 60 m, which can thus satisfy the requirement of large-scale and accurate mapping. Its five-day revisiting period also greatly enhances the availability of the images. In this study, we choose three water-sensitive bands for deep learning networks to learn the features of water bodies [47]. They are Short Wave Infra-Red (SWIR), Near Infra-Red (NIR), and Red with 20 m, 10 m and 10 m resolution, respectively. The 20 m SWIR thus is upsampled to 10 m by using the nearest neighboring algorithm.

To accelerate the labeling work for water bodies, the European Space Agency (ESA) WorldCover 10 m 2020 product is used as the supplementary data [48]. The ESA WorldCover 10 m 2020 product is a global land cover product with a resolution of 10 m, which is the same as Sentinel-2 images in spatial resolution. This product contains 11 land cover categories, including permanent water bodies, and the permanent water bodies are extracted to assist in labeling water bodies based on the Sentinel-2 images.

2.3. Water Body Extraction Network

The structure of water body extraction network is shown in Figure 2. The Swin Transformer [49] network which fully implements the self-attention mechanism in the computer vision (CV) field, is used to learn the water body feature. The attention mechanism selectively imitates the human behavior of paying attention to information, focusing attention on important information, and globally connecting contextual information to understand it. Unlike CNN, which only focuses on local information, the network based on an attentional mechanism [50] can obtain different attention of water body in the image and extract water body features at multiple levels from global to local. Thus, the transformer-based network can better capture the overall connection and difference between water body and other ground objects [51].

The Swin Transformer network has four modules for automatically extracting feature maps. They are the Swin Transformer Block, Patch Partition module, Linear Embedding module, and Patch Merging module. The Swin Transformer Block includes Window Multihead Self-Attention (W-MSA) and Shifted-Window Multihead Self-Attention (SW-MSA) to calculate the global attention. The W-MSA focuses on calculating self-attention

within the windows, also known as local self-attention, and the SW-MSA can interact with the information by shifting to obtain the global self-attention [52]. The Patch Partition module is used to convert the input images into different patches and change the minimum unit of the image from pixel to patch. The function of the Linear Embedding module is to transform the image into a one-dimensional vector. The Patch Merging module serves to downscale the feature maps and is able to form pyramid feature maps. The pyramid feature maps include different levels of context features and thus are good for more accurately extracting water bodies [53].

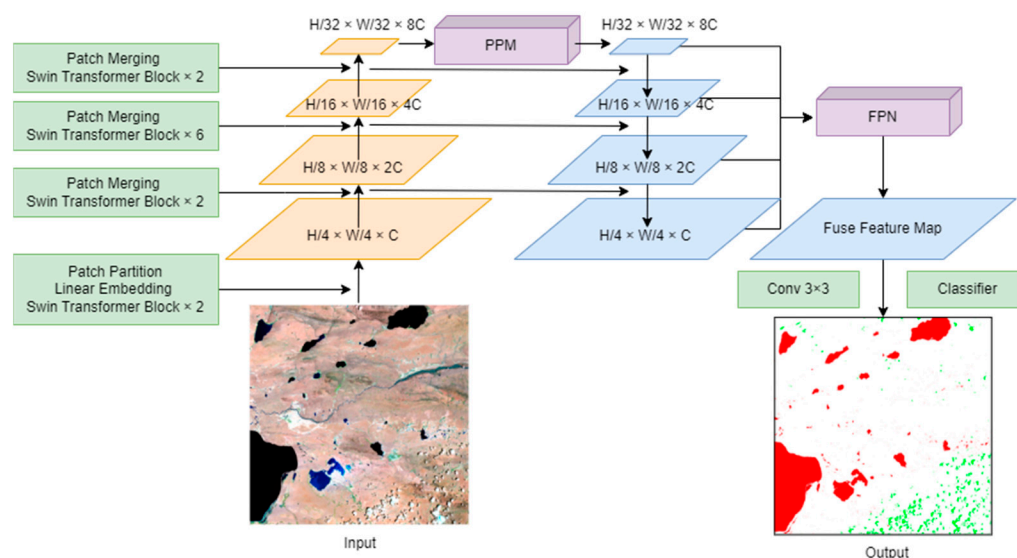


Figure 2. The structure of the water body extraction network.

To extract water bodies from the feature maps, Pyramid Pool Module (PPM) [54] is first used to obtain the scene-level feature, and Feature Pyramid Network (FPN) is used to combine the multi-scale object-level feature maps with the scene-level feature then obtain a fused feature map [55]. The fused feature map further goes through a 3×3 convolution, then a classifier is used to perform pixel-level classification.

2.4. Preparation for Training Data

The process of training data preparation is shown in Figure 3, which includes obtaining the SWIR-NIR-Red combined image data (SNR image data) from the Sentinel-2 imagery and obtaining ground truth data. The SNR image data is obtained by stacking the SWIR, NIR, Red bands from Sentinel-2 images.

To explore whether cloud shadow labels can improve the accuracy of water body extraction, two groups of training datasets are prepared: the group in which only water bodies are labeled, noted as the Water group, and the group in which both water bodies and cloud shadows are labeled, noted as the Water_Shadow group. The Water group includes the SNR image data and ground truth data that only water bodies are labeled, and the Water_Shadow group includes the SNR image data and ground truth data that both water bodies and cloud shadows are labeled.

To accelerate the labeling of water bodies, the permanent water bodies are extracted from the ESA World Cover 10 m 2020 product, then the water bodies are manually modified based on a selected Sentinel-2 image at a specific time. The modification includes the water boundary correction and addition of the missing water bodies.

During labeling the cloud shadows, no extra data, including the cloud or cloud shadow masks, was used as reference data. The cloud shadows are fully labeled manually based on the true color images from Sentinel-2 for ensuring the accuracy of cloud shadows.

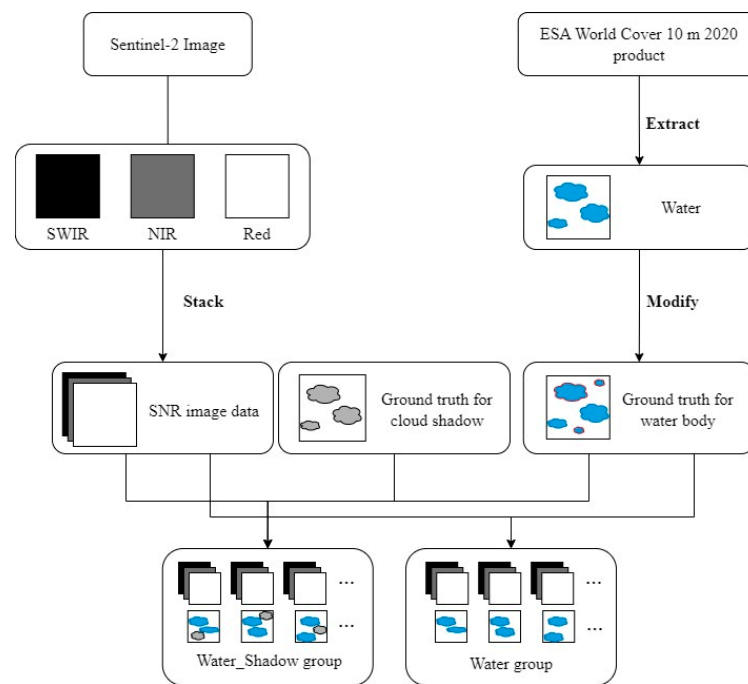


Figure 3. The process of preparation for training data.

3. Experiment and Results

3.1. Experiment Settings

Five Sentinel-2 images were selected to prepare for the training data, and the acquisition time of the five images is in the morning local time. A total of 8000 training samples were produced based on the Sentinel-2 images. The training samples are mainly located in the middle of the western Tibetan Plateau. Each of them is 272-pixel width and 272-pixel height due to the limitation of our GPU memory size. To explore how cloud shadows affect the accuracy of water body extraction, two training data groups are set up. One is the group that only water bodies are labeled, and the other is the group that both water bodies and cloud shadows are labeled. Each group includes two training datasets with 1% and 3% proportions of cloud shadows, respectively, as shown in Table 1. The proportion of water bodies in every training dataset is approximately 11% and the number of samples in all training datasets is 6400, which ensures the accuracy of water body extraction is not affected by the number of training samples and the proportion of water bodies.

Table 1. Training dataset settings.

Training Data Group	Training Dataset	Water Body		Cloud Shadow	
		Proportion	Labeled	Proportion	Labeled
Water group	Water_P1	11%	Yes	1%	No
	Water_P3	11%	Yes	3%	No
Water_Shadow group	Water_Shadow_P1	11%	Yes	1%	Yes
	Water_Shadow_P3	11%	Yes	3%	Yes

The validation dataset is made based on two additional Sentinel-2 images. The validation sets are also divided into a Water validation dataset and a Water_Shadow validation dataset. Each validation set contains 1012 samples with a size of 272×272 pixels, and the distribution of water bodies and cloud shadows are same. The only difference between the two validation datasets is that the cloud shadows are labeled in the Water_Shadow validation dataset.

The water body extraction network is trained for 230 epochs, and the top-three accuracy on the validation data are used to compare and analyze the performance of the network trained by different training data. The hardware environment in this study is an Intel(R) Xeon(R) W-2245 CPU @ 3.90 GHz with 16.0 GB RAM and an NVIDIA GeForce RTX3080 Ti GPU with 7424 CUDA cores. The initial learning rate is 0.000003, and it becomes 0.00006 by linearly changing small multiplicative factor after 10 epochs.

3.2. Metrics of Assessment

The metrics of assessment used in the experiment include the *Overall Accuracy (OA)*, *mean Intersection Over Union (mIoU)*, and *Kappa*.

The calculation formula of *OA* can be represented as:

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

OA is calculated by the confusion matrix, where *FP* denotes false-positive pixels, *FN* denotes false-negative pixels, *TP* denotes true-positive pixels, *TN* denotes true-negative pixels, and *N* denotes the number of pixels.

The calculation formula of the *mIoU* is:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FN + FP} \quad (2)$$

mIoU can also be calculated by a confusion matrix, where *k* is the number of classes.

The calculation formula of *Kappa* is:

$$Kappa = \frac{Po - Pe}{1 - Pe} \quad (3)$$

where *Po* is the *OA*, and *Pe* can be represented as:

$$Pe = \frac{a1 \times b1 + a2 \times b2 + \dots + ax \times bx}{N \times N} \quad (4)$$

where *a1*, *a2*... *ax* are the numbers of true values for every class and *b1*, *b2*..., *bx* are the numbers of prediction values for every class, and the *N* is the number of pixels.

3.3. Results

3.3.1. Accuracy Evaluation

Table 2 shows the top-three accuracy evaluation results of the water body extraction network trained by the four training datasets, and the average values are obtained by calculating the values based on the top-three accuracy evaluation results.

Table 2. Results of accuracy evaluation.

	Training Dataset	OA	mIoU	Kappa	Epoch
Top 1	Water_P1	0.9970	0.9734	0.9728	180
	Water_P3	0.9976	0.9778	0.9774	138
	Water_Shadow_P1	0.9891	0.9185	0.9459	61
	Water_Shadow_P3	0.9917	0.9429	0.9589	69
Top 2	Water_P1	0.9962	0.9664	0.9654	199
	Water_P3	0.9974	0.9766	0.9761	191
	Water_Shadow_P1	0.9886	0.9081	0.9429	60
	Water_Shadow_P3	0.9899	0.9317	0.9507	92

Table 2. Cont.

	Training Dataset	OA	mIoU	Kappa	Epoch
Top 3	Water_P1	0.9940	0.9488	0.9464	211
	Water_P3	0.9968	0.9714	0.9707	135
	Water_Shadow_P1	0.9876	0.9035	0.9383	90
	Water_Shadow_P3	0.9892	0.9252	0.9472	70
Average	Water_P1	0.9957	0.9629	0.9615	-
	Water_P3	0.9973	0.9753	0.9747	-
	Water_Shadow_P1	0.9884	0.91	0.9424	-
	Water_Shadow_P3	0.9903	0.9333	0.9523	-

In general, the accuracy on all four training datasets is high. The average OA values on the four training datasets are 0.9957, 0.9973, 0.9884 and 0.9903. The average mIoU values are 0.9629, 0.9753, 0.91 and 0.9333, respectively. The average Kappa values are 0.9615, 0.9747, 0.9424 and 0.9523, respectively. Thus, the overall performance of water body extraction using our water body extraction network is satisfactory. The training datasets in which only water bodies are labeled have higher values than the training datasets in which both water bodies and cloud shadows are labeled, which indicates that labeling cloud shadows will decrease the values of the metrics. When cloud shadows are not labeled, the “Water_P3” dataset has higher OA, mIoU and Kappa values than the “Water_P1” dataset in all results. When cloud shadows are labeled, the “Water_Shadow_P3” dataset has higher OA, mIoU and Kappa values than the “Water_Shadow_P1” dataset in all results. Thus, when the proportions of water body are similar, increasing the proportion of cloud shadows can improve the accuracy of the network. From the results, Water_P3 has the highest values. When cloud shadows are not labeled and the proportion is 3%, the network has the best accuracy.

For the epochs of training, the highest accuracy results in the four experimental sets appear in epochs 180, 199 and 211, epochs 138, 135 and 191, epochs 61, 60 and 90, and epochs 69, 92 and 70. This result shows that when cloud shadows are labeled, the network will converge faster.

3.3.2. Prediction Results

Five cases are summarized to illustrate the prediction results.

(1) Cloud shadows being misclassified as water bodies

Results for the Water Group: when cloud shadows are not labeled, the water body extraction network trained by the “Water_P1” dataset will be disturbed by cloud shadows. Some cloud shadows are misclassified as water bodies, especially the part in the middle of the cloud shadows with a large area, as shown in Figure 4b. However, when the proportion of unlabeled cloud shadows is 3%, the phenomenon of cloud shadows being misclassified as water bodies is significantly reduced, and the interference of cloud shadows is well eliminated (Figure 4c).

Results for the Water_Shadow Group: when cloud shadows are labeled, the water body extraction network can predict most boundaries of cloud shadows, as shown in Figure 4d. The phenomenon of cloud shadows being misclassified as water bodies is less when the proportion is 3% than when the proportion is 1% (Figure 4e). However, compared with experimental sets in which cloud shadows are not labeled, the areas in the middle of large cloud shadows are still misclassified as water bodies. Therefore, labeling cloud shadows does not make cloud shadows less misclassified than adding cloud shadows but without labeling.

(2) Precision of water boundaries

Results for the Water Group: when cloud shadows are not labeled and the proportion is 1%, there are some noise points on the boundary of lakes, as shown in Figure 5b. However, when the proportion of cloud shadows increases to 3%, the noise points on the boundary are eliminated, and the extracted lake boundaries are more accurate (Figure 5c). Thus, increasing the proportion of cloud shadows in the training dataset helps to extract more accurate water boundaries.

Results for the Water_Shadow Group: when cloud shadows are labeled, regardless of whether the proportion of cloud shadows is 1% or 3%, the results are identical to those for the Water_P3 dataset, as shown in Figure 5d,e. The water boundaries are also accurately predicted. Thus, when cloud shadows are labeled, the water boundaries are more accurately predicted, and the proportion does not affect the accuracy of the boundaries.

(3) Water bodies being missed during extraction

Results for the Water Group: when cloud shadows are not labeled, the water body extraction network trained by 1% cloud shadows and 3% cloud shadows in the training dataset can predict most water bodies, as shown in Figure 6b,c, even the small lakes that cover several pixels. This result indicates that when cloud shadows are not labeled, the proportion of cloud shadows hardly affects the number of predicted water bodies.

Results for the Water_Shadow Group: when cloud shadows are labeled, the number of predicted water bodies decreases compared with those of the Water group, as shown in Figure 6d,e. Thus, labeling cloud shadows reduces the number of predicted water bodies.

When the proportion of labeled cloud shadows is 3%, the number of predicted water bodies by the network is less than that predicted by the network when the proportion of cloud shadows is 1%. Thus, when cloud shadows are labeled, the proportion of cloud shadows in the training dataset is higher, and the number of predicted water bodies is fewer.

(4) The middle part of islands being misclassified as water body

Results for the Water Group: when cloud shadows are not labeled and the proportion is 1%, the middle part of the islands in the lakes is misclassified as water body, as shown in Figure 7b. However, this phenomenon disappears when the proportion of cloud shadows increases to 3% (Figure 7c). Thus, when cloud shadows are not labeled, increasing the proportion of cloud shadows in the training dataset can make the water body extraction network better predict the islands.

Results for the Water_Shadow Group: when cloud shadows are labeled, regardless of whether the proportion of cloud shadows is 1% or 3%, the prediction result of the islands is identical to that of the Water_P3 dataset, as shown in Figure 7d,e. The middle part of islands is not misclassified as water body, and the prediction result is more accurate. Thus, labeling cloud shadows can make the water body extraction network better distinguish the islands from the water bodies.

(5) Cloud shadow extraction results

When comparing our extraction results of cloud shadows with those of Fmask, the boundaries of our cloud shadow results are more refined than the boundaries extracted by Fmask. The cloud shadow boundaries extracted by Fmask are coarser than the true boundaries, as shown in Figure 8b.

When the proportion of cloud shadows is 1% in the training dataset, our deep learning network can predict most cloud shadow boundaries. However, the extracted cloud shadows are mostly fragmented and incomplete, as shown in Figure 8c. When the proportion of cloud shadows is increased to 3%, the boundaries of cloud shadows are more accurate and complete than when the proportion is 1%, as shown in Figure 8d. Thus, increasing the proportion of cloud shadows can improve the accuracy and completeness of extracted cloud shadows.

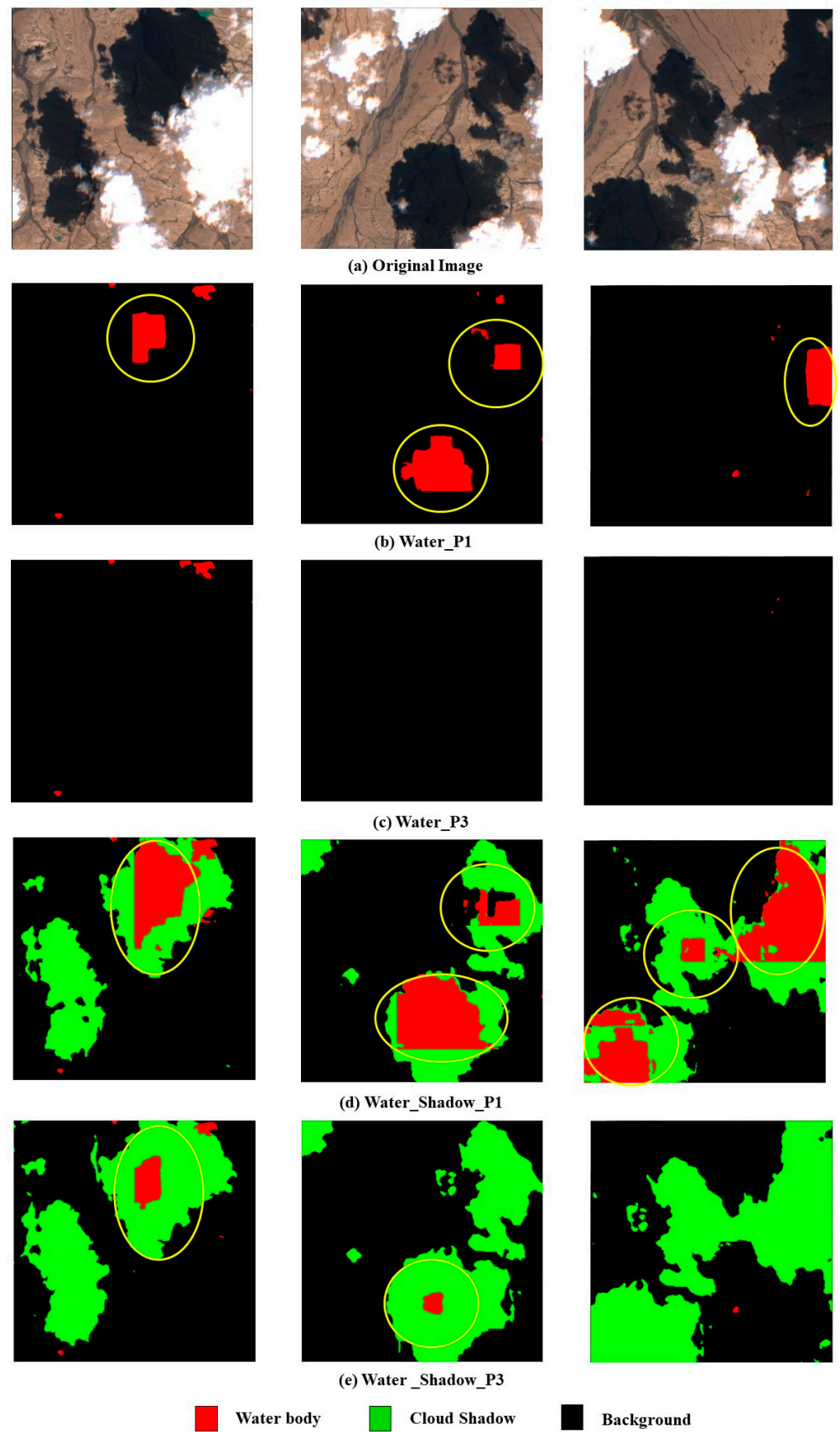


Figure 4. Results of cloud shadows being misclassified as water bodies (The areas circled in yellow represent the cloud shadows being misclassified as water bodies).

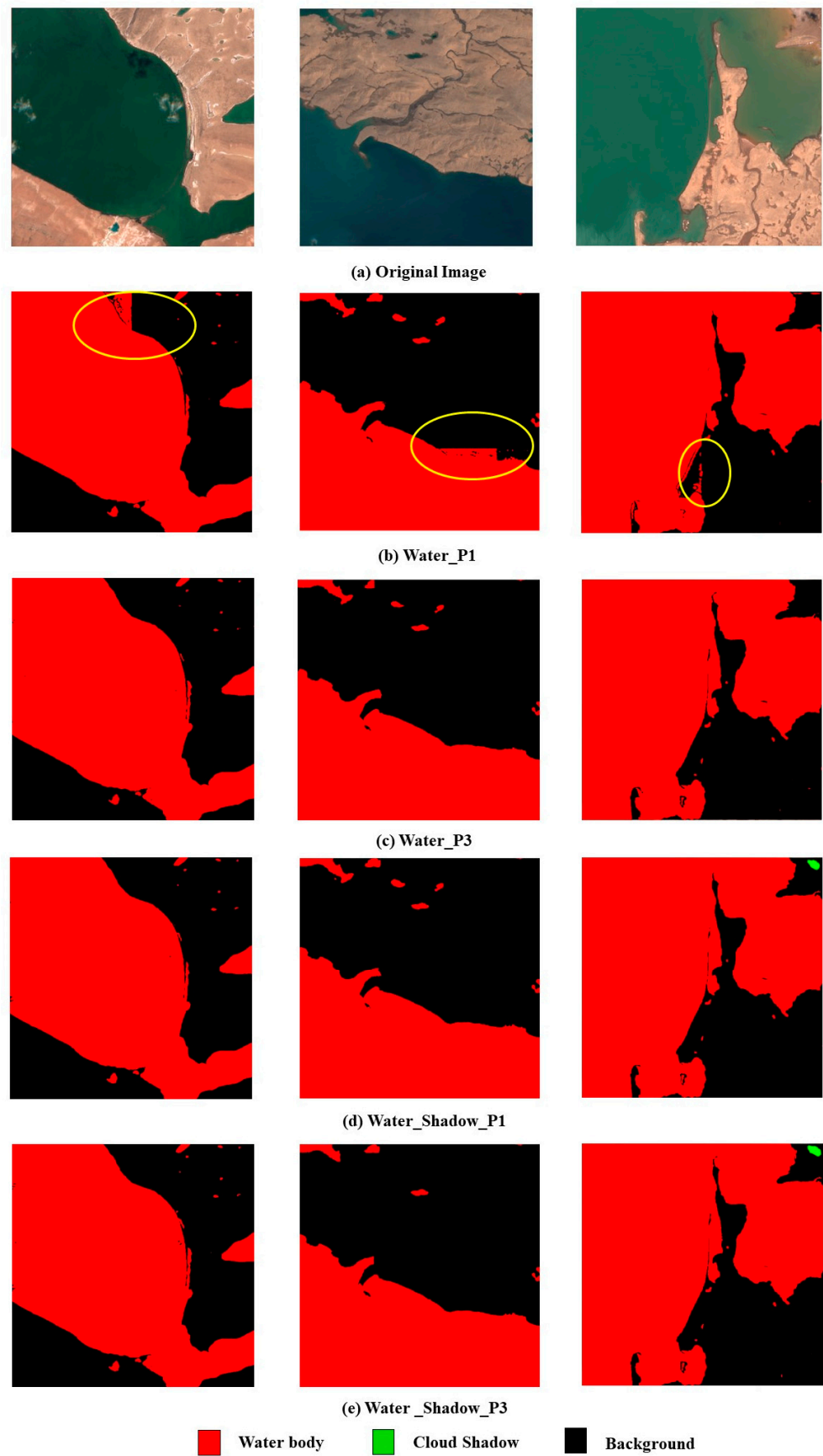


Figure 5. Results of water boundary extraction (The areas circled in yellow represent the noise points on the boundaries of water bodies).

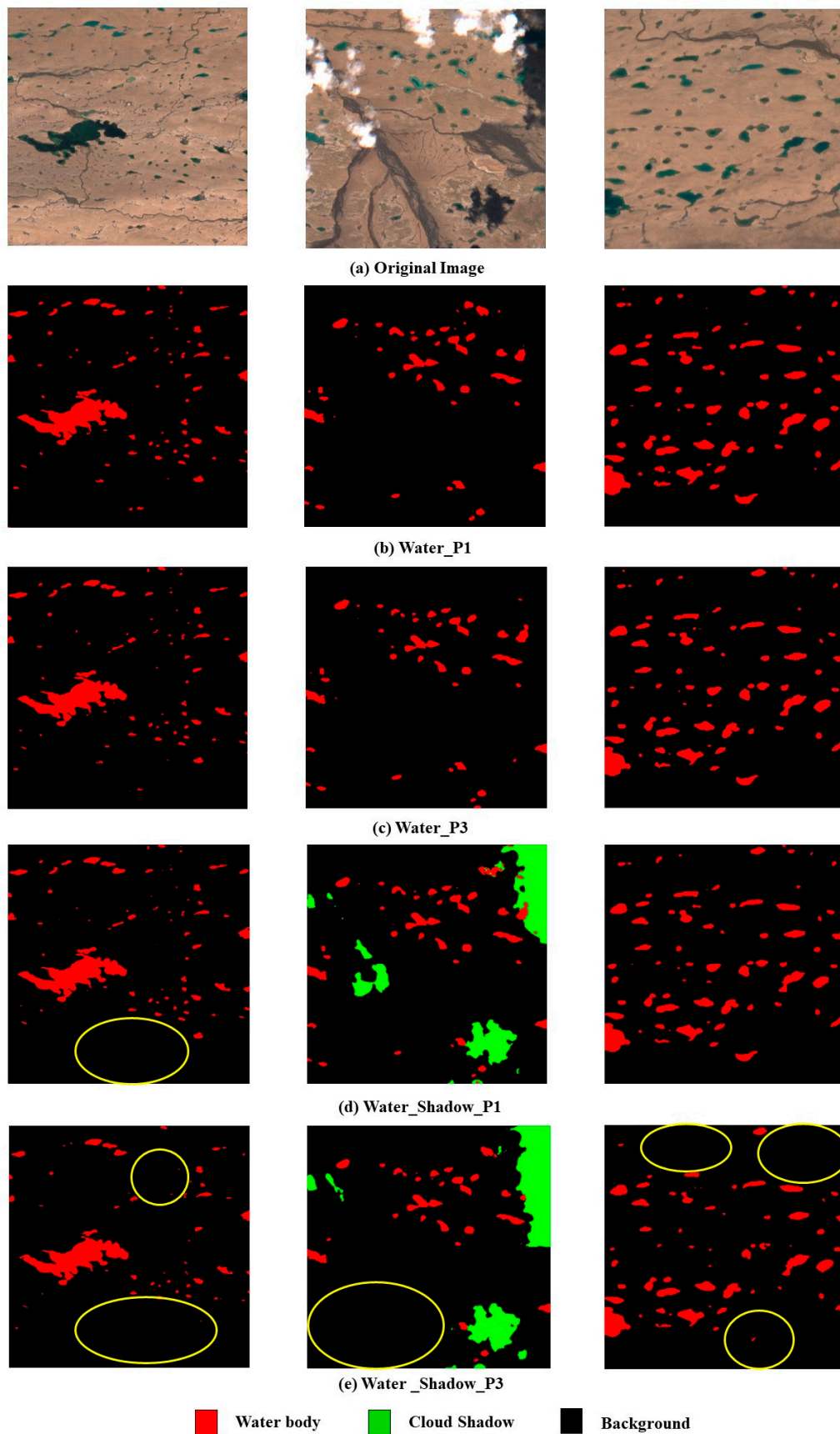


Figure 6. Results of water bodies being missed (The areas circled in yellow represent the water bodies being missed).

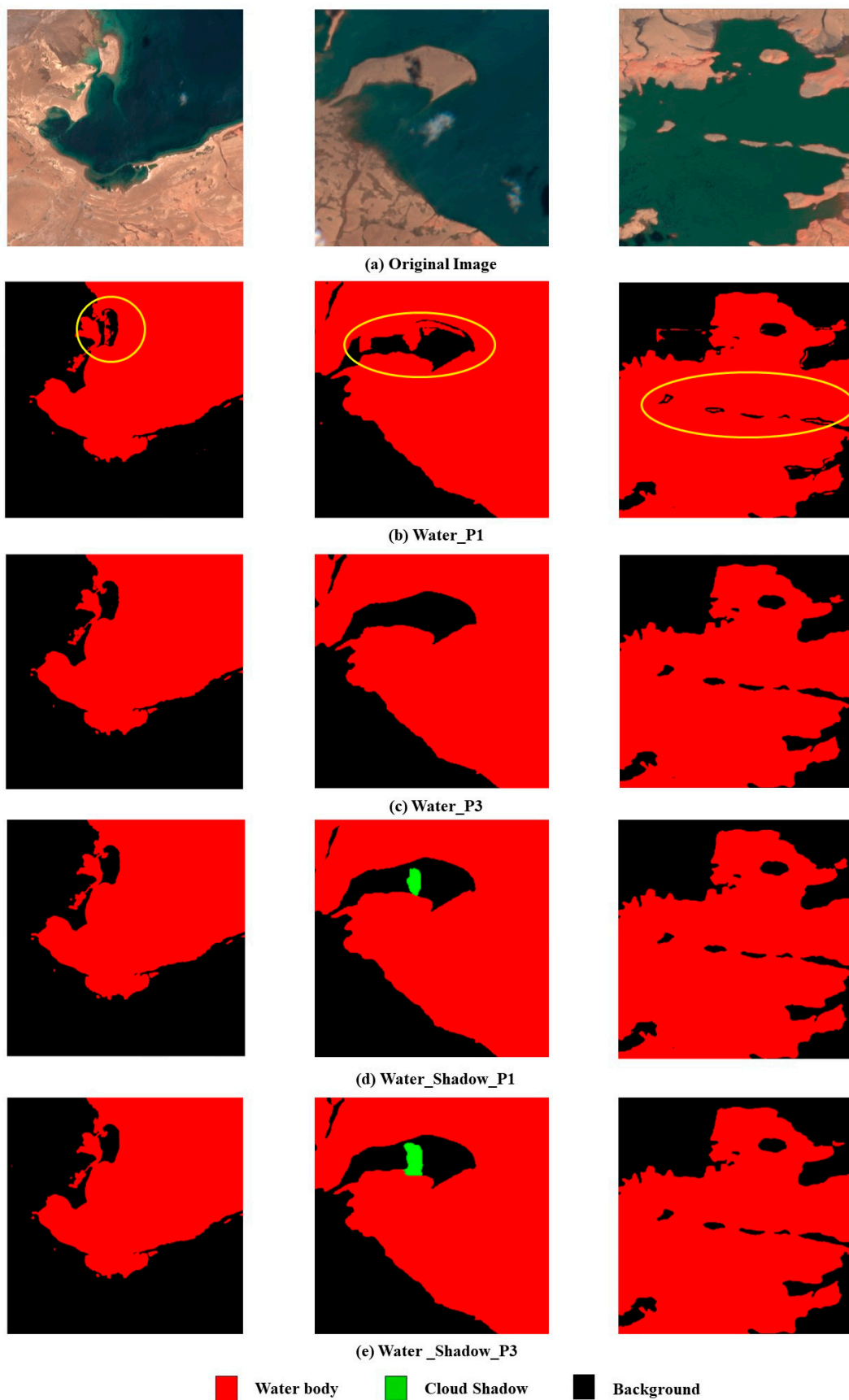


Figure 7. Results of the middle part of islands being misclassified as water body (The areas circled in yellow represent the middle part of islands being misclassified as water body).

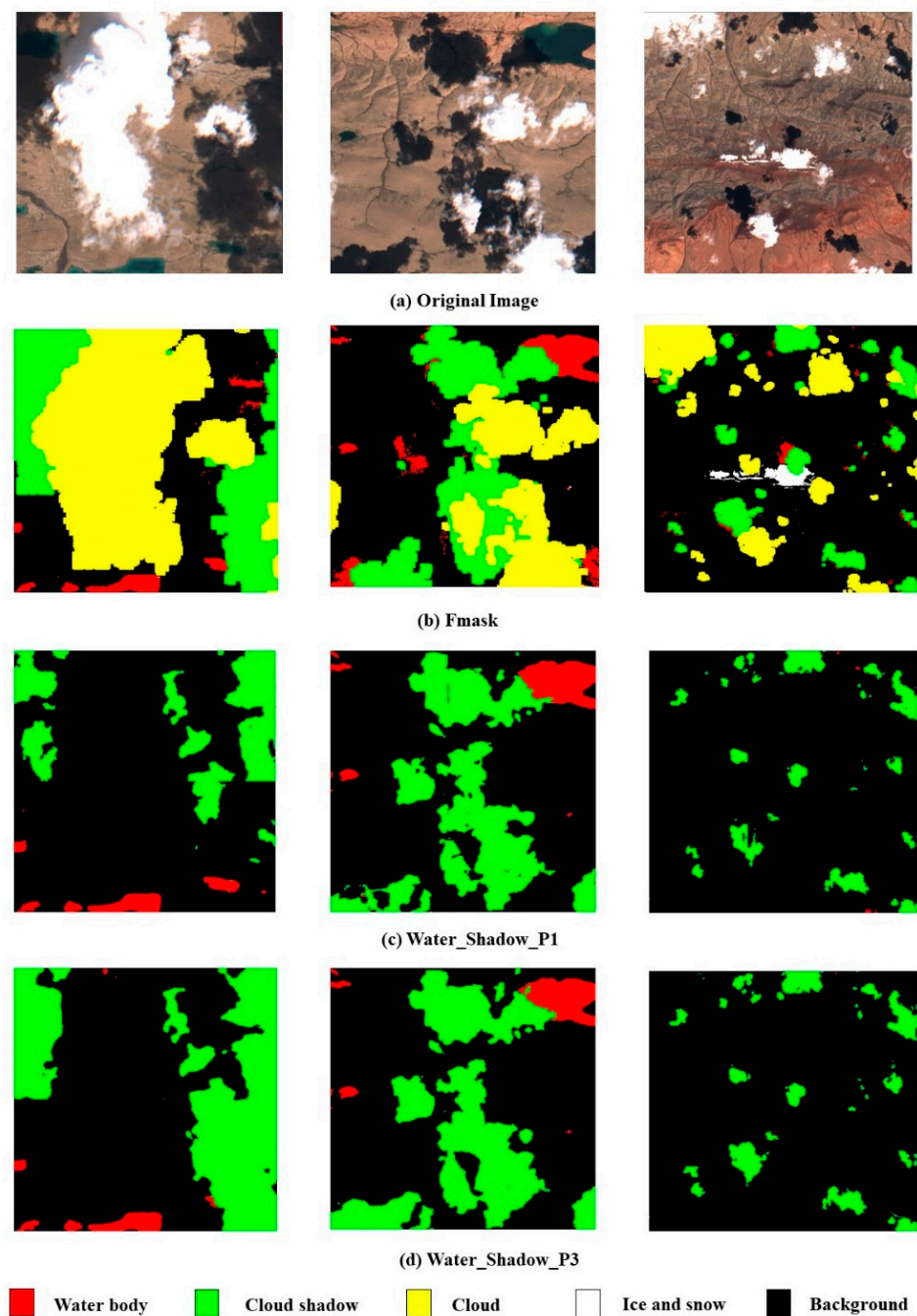


Figure 8. Results of cloud shadow extraction.

4. Discussion

While the existing studies pay more attention to the improvement or innovation of deep learning models for accurate water body extraction, this study highlights the function of negative samples. It reveals that the negative samples play an important role in achieving good or satisfactory accuracy, and confirms that cloud shadow negative samples do affect the accuracy of water body extraction from remote sensing images. More specifically, the study demonstrates an effective method based on the negative samples to address the misclassification of cloud shadows thereby improving the accuracy of water body extraction. Based on the experiment of this study, it can be seen that even if the negative samples change slightly (the proportion increases from 1% to 3% in this study), the prediction result of water bodies changes significantly. A more detailed discussion about the cloud shadow negative samples being categorized into cloud shadow misclassification, the precision of

the water boundaries, small water body extraction, as well as the precision of the extracted cloud shadows follows:

Cloud shadow misclassification: the cloud shadow negative samples help to address the issue of cloud shadow misclassification in water body extraction. When 1% cloud shadow negative samples are introduced, fewer cloud shadows are misclassified as water bodies despite that there are still a few cloud shadows misclassified. When the proportion of the cloud shadows reaches 3% and even if the cloud shadows are not labeled, the water body extraction network is able to handle the misclassification of cloud shadows. In a word, a higher proportion of cloud shadow negative samples in the training datasets corresponds to a stronger ability of the network to eliminate the interference of cloud shadows. In addition, labeling the cloud shadows cannot significantly further reduce the cloud shadow misclassification. We believe the reason is that prediction of both water bodies and cloud shadows is more challenging than only predicting water bodies for the network. Thus, the accuracy of water body extraction could be even slightly low when extracting both water bodies and cloud shadows at the same time.

Precision of the water boundaries: the experiment shows that when the proportion of cloud shadows is 1%, labeling cloud shadows helps to extract more accurate water boundaries, and when the proportion of cloud shadows increases to 3%, both labeling and not labeling cloud shadows can further improve the precision of the water boundaries. We believe that when the cloud shadow negative samples in the training samples reach a certain proportion (3% in our experiment), the water body extraction network learned more features of cloud shadows and thus can accurately distinguish water bodies from cloud shadows thereby extracting more accurate water boundaries.

Small water body extraction: the experiments show that labeling the cloud shadows causes the small water bodies to be missed in water body extraction. The feature of small water bodies is obscure relative to large water bodies, and they are more easily missed when the network conducts more challenging tasks such as multi-classification. Therefore, if the proportion of cloud shadows is sufficiently high, it is more conducive to water body extraction without labeling cloud shadows.

Precision of the extracted cloud shadows: with the training datasets in which cloud shadows are labeled, the experiment proved that the boundary of cloud shadows identified by the proposed transformer-based network is more accurate and refined than the ones identified by Fmask [40]. Although, the proportion of cloud shadows in the training dataset is only 3% and not high, the network can extract both water bodies and cloud shadows at the same time. This demonstrates the great potential of the novel transformer-based deep learning networks applying in remote sensing field.

Lastly, although the network achieves the best performance on water body extraction with 3% cloud shadow negative samples, the proportion of 3% is only applicable to cloud shadow samples, and meets 6400 samples and every image size of 272×272 pixels. We believe that the method of this study could be applied on other negative samples such as mountain shadows and building shadows for eliminating the interference of them in water body extraction. When applying other negative samples, the 3% may not be an appropriate proportion of negative samples. However, the appropriate proportion of other negative samples can be explored based on the method of this study.

5. Conclusions

Deep learning is one of the most effective approaches to extract water bodies from remote sensing images, and making samples for training the deep learning networks is indispensable. This study pays close attention to the negative samples to explore the impact of the cloud shadow negative samples on the accuracy of water body extraction. The western Tibetan Plateau is the study area for producing training data and validating prediction results. A water body extraction based on a novel vision transformer network was built, and the network was trained with the datasets that contain different proportions of cloud shadows. It was found that the training datasets containing a certain proportion

of cloud shadows could help the water body extraction network to better distinguish water bodies and cloud shadows, and make the extracted water bodies more accurate. With 3% cloud shadow negative samples in this study, the water body extraction network can well address the issue of cloud shadows being misclassified as the water bodies, and the evaluation results over the validation dataset achieve *OA* of 0.9973, *mIoU* of 0.9753, and *Kappa* of 0.9747. The network was also trained with the datasets in which cloud shadows were labeled to investigate whether labeling the cloud shadows is helpful. It was found that labeling the cloud shadows in the training data is unnecessary. In addition, this study also reveals the cloud shadow boundaries predicted by our network are more accurate than the ones using the Fmask method. In the future, we will conduct more studies on other negative samples, such as mountain shadows and building shadows, to investigate whether they have the similar results to this study, and what the appropriate proportion of them is. In addition, we may develop other metrics to quantify the spectral and morphological features of negative samples and better explain the function of negative samples in improving the performance of deep learning networks.

Author Contributions: J.S.: conceptualization, methodology, software, resources, funding acquisition, writing—review and editing, supervision. X.Y.: methodology, validation, data curation, formal analysis, visualization, writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding: The National Key Research and Development Program of China (2022YFF0711602, 2021YFE0117800), the 14th Five-year Informatization Plan of Chinese Academy of Sciences (CAS-WX2021SF-0106), and the National Data Sharing Infrastructure of Earth System Science (<http://www.geodata.cn/> accessed on 16 July 2022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data generated and analyzed during this study are available from the corresponding author by request.

Acknowledgments: We appreciate the detailed comments from the editor and the anonymous reviewers.

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

References

1. Zhang, Z.; Prinet, V.; Ma, S. Water Body Extraction from Multi-Source Satellite Images. In Proceedings of the IGARSS 2003—2003 IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003; Volume 6, pp. 3970–3972. [[CrossRef](#)]
2. Li, J.; Ma, R.; Cao, Z.; Xue, K.; Xiong, J.; Hu, M.; Feng, X. Satellite Detection of Surface Water Extent: A Review of Methodology. *Water* **2022**, *14*, 1148. [[CrossRef](#)]
3. Shrestha, R.; Di, L. Land/Water Detection and delineation with Landsat Data Using Matlab/ENVI. In Proceedings of the 2013 Second International Conference on Agro-Geoinformatics, Fairfax, VA, USA, 12–16 August 2013; pp. 211–214. [[CrossRef](#)]
4. Sit, M.A.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* **2020**, *82*, 2635–2670. [[CrossRef](#)] [[PubMed](#)]
5. Li, M.; Hong, L.; Guo, J.; Zhu, A. Automated Extraction of Lake Water Bodies in Complex Geographical Environments by Fusing Sentinel-1/2 Data. *Water* **2022**, *14*, 30. [[CrossRef](#)]
6. Li, J.; Meng, Y.; Li, Y.; Cui, Q.; Yang, X.; Tao, C.; Wang, Z.; Li, L.; Zhang, W. Accurate water extraction using remote sensing imagery based on normalized difference water index and unsupervised deep learning. *J. Hydrol.* **2022**, *612*, 128202. [[CrossRef](#)]
7. Shen, C. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour. Res.* **2018**, *54*, 8558–8593. [[CrossRef](#)]
8. Yu, L.; Zhang, R.; Tian, S.; Yang, L.; Lv, Y. Deep Multi-Feature Learning for Water Body Extraction from Landsat Imagery. *Autom. Control. Comput. Sci.* **2018**, *52*, 517–527. [[CrossRef](#)]
9. Wang, Z.; Gao, X.; Zhang, Y.; Zhao, G. MSLWENet: A Novel Deep Learning Network for Lake Water Body Extraction of Google Remote Sensing Images. *Remote Sens.* **2020**, *12*, 4140. [[CrossRef](#)]
10. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; Guan, L.; Fang, H. Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7422–7434. [[CrossRef](#)]

11. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Honolulu, HI, USA, 21–26 July 2017; Volume 39, pp. 640–651.
12. Li, L.; Yan, Z.; Shen, Q.; Cheng, G.; Gao, L.; Zhang, B. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. *Remote Sens.* **2019**, *11*, 1162. [[CrossRef](#)]
13. Zhang, J.; Xing, M.; Sun, G.-C.; Chen, J.; Li, M.; Hu, Y.; Bao, Z. Water Body Detection in High-Resolution SAR Images With Cascaded Fully-Convolutional Network and Variable Focal Loss. *IEEE Trans. Geosci. Remote. Sens.* **2021**, *59*, 316–332. [[CrossRef](#)]
14. Song, A.; Kim, Y.; Kim, Y. Change detection of surface water in remote sensing images based on fully convolutional network. *J. Coast. Res.* **2019**, *91*, 426–430. [[CrossRef](#)]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science, Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; Volume 9351. [[CrossRef](#)]
16. Ch, A.; Ch, R.; Gadamssetty, S.; Iwendi, C.; Gadekallu, T.R.; Dhaou, I.B. ECDSA-Based Water Bodies Prediction from Satellite Images with UNet. *Water* **2022**, *14*, 2234. [[CrossRef](#)]
17. An, S.; Rui, X. A High-Precision Water Body Extraction Method Based on Improved Lightweight U-Net. *Remote Sens.* **2022**, *14*, 4127. [[CrossRef](#)]
18. Jiang, C.; Zhang, H.; Wang, C.; Ge, J.; Wu, F. Water Surface Mapping from Sentinel-1 Imagery Based on Attention-UNet3+: A Case Study of Poyang Lake Region. *Remote Sens.* **2022**, *14*, 4708. [[CrossRef](#)]
19. Ge, C.; Xie, W.; Meng, L. Extracting Lakes and Reservoirs from GF-1 Satellite Imagery Over China Using Improved U-Net. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1504105. [[CrossRef](#)]
20. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
21. James, T.; Schillaci, C.; Lipani, A. Convolutional neural networks for water segmentation using sentinel-2 red, green, blue (RGB) composites and derived spectral indices. *Int. J. Remote Sens.* **2021**, *42*, 5338–5365. [[CrossRef](#)]
22. Wu, H.; Song, H.; Huang, J.; Zhong, H.; Zhan, R.; Teng, X.; Qiu, Z.; He, M.; Cao, J. Flood Detection in Dual-Polarization SAR Images Based on Multi-Scale Deeplab Model. *Remote Sens.* **2022**, *14*, 5181. [[CrossRef](#)]
23. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
24. Balaska, V.; Bampis, L.; Kansizoglou, I.; Gasteratos, A. Enhancing satellite semantic maps with ground-level imagery. *Robot. Auton. Syst.* **2021**, *139*, 103760. [[CrossRef](#)]
25. Weng, L.; Xu, Y.; Xia, M.; Zhang, Y.; Liu, J.; Xu, Y. Water Areas Segmentation from Remote Sensing Images Using a Separable Residual SegNet Network. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 256. [[CrossRef](#)]
26. Pu, F.; Ding, C.; Chao, Z.; Yu, Y.; Xu, X. Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1674. [[CrossRef](#)]
27. Wang, Y.; Li, Z.; Zeng, C.; Xia, G.-S.; Shen, H. An Urban Water Extraction Method Combining Deep Learning and Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 769–782. [[CrossRef](#)]
28. Guo, H.; He, G.; Jiang, W.; Yin, R.; Yan, L.; Leng, W. A Multi-Scale Water Extraction Convolutional Neural Network (MWEN) Method for GaoFen-1 Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 189. [[CrossRef](#)]
29. Luo, X.; Tong, X.; Hu, Z. An applicable and automatic method for earth surface water mapping based on multispectral images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102472. [[CrossRef](#)]
30. Fei, J.; Liu, J.; Ke, L.; Wang, W.; Wu, P.; Zhou, Y. A deep learning-based method for mapping alpine intermittent rivers and ephemeral streams of the Tibetan Plateau from Sentinel-1 time series and DEMs. *Remote Sens. Environ.* **2022**, *282*, 113271. [[CrossRef](#)]
31. Li, M.; Wu, P.; Wang, B.; Park, H.; Hui, Y.; Yanlan, W. A Deep Learning Method of Water Body Extraction From High Resolution Remote Sensing Images with Multisensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3120–3132. [[CrossRef](#)]
32. Chen, F. Comparing Methods for Segmenting Supra-Glacial Lakes and Surface Features in the Mount Everest Region of the Himalayas Using Chinese GaoFen-3 SAR Images. *Remote Sens.* **2021**, *13*, 2429. [[CrossRef](#)]
33. Yang, F.; Guo, J.; Tan, H.; Wang, J. Automated Extraction of Urban Water Bodies from ZY-3 Multi-Spectral Imagery. *Water* **2017**, *9*, 144. [[CrossRef](#)]
34. Dirscherl, M.; Dietz, A.J.; Kneisel, C.; Kuenzer, C. A Novel Method for Automated Supraglacial Lake Mapping in Antarctica Using Sentinel-1 SAR Imagery and Deep Learning. *Remote Sens.* **2021**, *13*, 197. [[CrossRef](#)]
35. Jiang, D.; Li, X.; Zhang, K.; Marinsek, S.; Hong, W.; Wu, Y. Automatic Supraglacial Lake Extraction in Greenland Using Sentinel-1 SAR Images and Attention-Based U-Net. *Remote Sens.* **2022**, *14*, 4998. [[CrossRef](#)]
36. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
37. Kansizoglou, I.; Bampis, L.; Gasteratos, A. Deep feature space: A geometrical perspective. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 6823–6838. [[CrossRef](#)] [[PubMed](#)]
38. Wang, G.; Wu, M.; Wei, X.; Song, H. Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 795. [[CrossRef](#)]

39. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [[CrossRef](#)]
40. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [[CrossRef](#)]
41. Zhu, Z.; Woodcock, C.E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* **2014**, *152*, 217–234. [[CrossRef](#)]
42. Kang, S.; Xu, Y.; You, Q.; Flügel, W.-A.; Pepin, N.; Yao, T. Review of climate and cryospheric change in the Tibetan Plateau. *Environ. Res. Lett.* **2010**, *5*, 015101. [[CrossRef](#)]
43. Zhang, G.; Xie, H.; Kang, S.; Yi, D.; Ackley, S.F. Monitoring lake level changes on the Tibetan Plateau using ICESat altimetry data (2003–2009). *Remote Sens. Environ.* **2011**, *115*, 1733–1742. [[CrossRef](#)]
44. Song, C.; Huang, B.; Ke, L.; Richards, K.S. Remote sensing of alpine lake water environment changes on the Tibetan Plateau and surroundings: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 26–37. [[CrossRef](#)]
45. Zhang, G.; Li, J.; Zheng, G. Lake-area mapping in the Tibetan Plateau: An evaluation of data and methods. *Int. J. Remote Sens.* **2017**, *38*, 742–772. [[CrossRef](#)]
46. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
47. Zhang, M.; Wang, X.; Shi, C.; Yan, D. Automated Glacier Extraction Index by Optimization of Red/SWIR and NIR /SWIR Ratio Index for Glacier Mapping Using Landsat Imagery. *Water* **2019**, *11*, 1223. [[CrossRef](#)]
48. Zanaga, D.; van de Kerchove, R.; de Keersmaecker, W.S.N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; Cartus, O.; et al. *ESA WorldCover 10 m 2020 v100*; The European Space Agency: Paris, France, 2021. [[CrossRef](#)]
49. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [[CrossRef](#)]
50. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; Volume 30. [[CrossRef](#)]
51. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [[CrossRef](#)]
52. Vaswani, A.; Ramachandran, P.; Srinivas, A.; Parmar, N.; Hechtman, B.; Shlens, J. Scaling local self-attention for parameter efficient visual backbones. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; p. 12889. [[CrossRef](#)]
53. 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](#)]
54. Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; Sun, J. Unified Perceptual Parsing for Scene Understanding. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 432–448. [[CrossRef](#)]
55. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [[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.