remote sensing

MDPI

Article

Terrain Shadow Interference Reduction for Water Surface Extraction in the Hindu Kush Himalaya Using a Transformer-Based Network

Xiangbing Yan [1,2,†] and Jia Song [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; 2017301110181@whu.edu.cn
[2] School of Resource and Environmental Science, Wuhan University, Wuhan 430072, China
[3] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
* Correspondence: songj@igsnrr.ac.cn
† These authors contributed equally to this work.

Abstract: Water is the basis for human survival and growth, and it holds great importance for ecological and environmental protection. The Hindu Kush Himalaya (HKH) is known as the "Water Tower of Asia", where water influences changes in the global water cycle and ecosystem. It is thus very important to efficiently measure the status of water in this region and to monitor its changes; with the development of satellite-borne sensors, water surface extraction based on remote sensing images has become an important method through which to do so, and one of the most advanced and accurate methods for water surface extraction involves the use of deep learning networks. We designed a network based on the state-of-the-art Vision Transformer to automatically extract the water surface in the HKH region; however, in this region, terrain shadows are often misclassified as water surfaces during extraction due to their spectral similarity. Therefore, we adjusted the training dataset in different ways to improve the accuracy of water surface extraction and explored whether these methods help to reduce the interference of terrain shadows. Our experimental results show that, based on the designed network, adding terrain shadow samples can significantly enhance the accuracy of water surface extraction in high mountainous areas, such as the HKH region, while adding terrain data does not reduce the interference from terrain shadows. We obtained the water surface extraction results in the HKH region in 2021, with the network and training datasets containing both water surface and terrain shadows. By comparing these results with the data products of Global Surface Water, it was shown that our water surface extraction results are highly accurate and the extracted water surface boundaries are finer, which strongly confirmed the applicability and advantages of the proposed water surface extraction approach in a wide range of complex surface environments.

Keywords: deep learning; terrain shadow; hydrology; high mountain; Transformer; semantic segmentation

Citation: Yan, X.; Song, J. Terrain Shadow Interference Reduction for Water Surface Extraction in the Hindu Kush Himalaya Using a Transformer-Based Network. *Remote Sens.* **2024**, *16*, 2032. https://doi.org/10.3390/rs16112032

Academic Editors: Mohib Ullah, Sultan Daud Khan and Habib Ullah

Received: 26 March 2024
Revised: 31 May 2024
Accepted: 1 June 2024
Published: 5 June 2024

1. Introduction

Water has been described as the "catalyst of life on Earth and the source of all things". As an indispensable source of energy for human society, water influences human economic prosperity and productive development [1,2]. The Hindu Kush Himalaya (HKH), also known as the "Water Tower of Asia", supplies water to 1.3 billion people and affects the ecosystems of river basins both within and beyond this region [3,4]. It is of great significance to efficiently obtain the spatial distribution of water surface in this region in order to monitor its status and changes [5,6]. The HKH is one of the largest mountain systems in the world; its high mountains, intermountain valleys, and plateaus result in

complex terrain and undulating topography that make terrain shadow interference a major challenge for water surface extraction in this region [7].

Over several years of research, many different types of water surface extraction methods have been developed, including single- and multi-band spectral analysis, water indexing, shallow machine learning, and deep learning methods [8]. Both single- and multi-band spectral analysis methods, as well as water index methods [9,10], mainly achieve water surface extraction based on human a priori cognitive knowledge, but they can be considered as water surface extraction methods based on artificially created features [11]. On the other hand, the shallow machine learning and deep learning methods allow machines to automatically learn water features based on samples [12,13]. Moreover, the deep learning-based method is an end-to-end water surface extraction method with higher accuracy and large-scale applicability that has been developed with the recent breakthroughs in artificial intelligence (AI) [14–16]. The advantage of the deep learning-based water surface extraction method is that the artificial neural network used in deep learning facilitates consideration of the complex nonlinear features of the heterogeneous surface [17–19]. Therefore, the models based on deep neural networks are more robust than the shallow machine learning and traditional methods based on spectral analysis and water indexes [20–22].

However, shadows, such as those from terrain, clouds, and other phenomena, are often major distractions during water surface extraction due to their similar spectral features to water surfaces. The HKH region is the most undulating and extensive mountainous region in Asia and, indeed, the world; therefore, the accuracy of existing water surface products, such as the European Commission's Joint Research Center's Global Surface Water (JRC GSW) [23], is often not as good as it could be for this region due to strong interference from terrain shadows. In previous studies, two main approaches have been investigated to avoid or reduce terrain shadow interference: One approach involves the utilization of Digital Surface Model (DSM) or Digital Elevation Model (DEM) data [24–26] to obtain various terrain factor information, which is then used as terrain-related prior knowledge in order to remove the terrain shadows that are misclassified as water surfaces [27,28], an approach that obviously depends on the accuracy of the terrain data and the reliability of terrain-related prior knowledge [29,30]. The other approach does not use any terrain data, but only uses spectral images to improve water surface extraction in the mountainous area based on the band math calculation algorithm [31].

The issue of water surface extraction being disturbed by terrain shadows is not well resolved due to manual modeling and potentially limited a priori knowledge of the terrain [23,32]. Considering the data-driven AI models [33], especially the big models based on the novel Transformer architecture, which show excellent performance in both the computer vision and remote sensing communities [34–37], this study targets big data-driven AI models by introducing samples of terrain shadows in high mountainous areas, exploring whether these models can effectively distinguish water surfaces from terrain shadows only based on the spectral information and whether the terrain data, such as DSM or DEM, are highly useful in distinguishing water surfaces and terrain shadows. Therefore, in this study, we developed an end-to-end Transformer-based AI model for water surface extraction, and the remote sensing image samples containing terrain shadows in the HKH region were specifically prepared with the aim of investigating the effect of adding terrain shadow samples on the interference of these shadows during water surface extraction. Furthermore, we also introduced DSM data and prepared specific samples containing the relevant terrain features in order to investigate whether terrain data significantly help to reduce the interference of terrain shadows during water surface extraction with the Transformer-based AI models. Finally, we extracted surface water distribution data over the whole HKH region, using the model that demonstrated the best performance in our experiment, and compared them with the JRC GSW data to evaluate the accuracy of our extraction results and to verify the robustness of the proposed method to reduce the interference of terrain shadows in particularly large and high mountainous areas.

2. Study Area and Data Sources

2.1. Study Area

The entire Hindu Kush Himalaya (HKH) region was adopted as the study area for our research (Figure 1). This area is located at 60.8539°E~105.0447°E, 15.9579°N~39.3187°N. Many large and important lakes and rivers are distributed in this area, such as Lake Qinghai, Lake Selincuo, Lake Namtso, and the Yangtze and Yellow Rivers [38], thus affecting the water circulation in the region and worldwide [39,40]. The HKH region also includes many globally significant mountains, including the Himalaya, Pamir, Tianshan, and Hengduan Mountains, among others [41]. These mountains contribute to the unique climatic conditions and rich biodiversity of the region (Figure 2) [42,43]; however, they also pose a significant terrain shadow interference problem for water surface extraction from remote sensing images. Therefore, the selected HKH region is appropriate for studying the influence of terrain shadows during water surface extraction based on remote sensing images. Additionally, the HKH region covers more than 4.3 million km^2 and spans eight countries, which is ideal for demonstrating that the Transformer-based method for reducing terrain shadow interference during water surface extraction can be applied to a large, high, mountainous area.

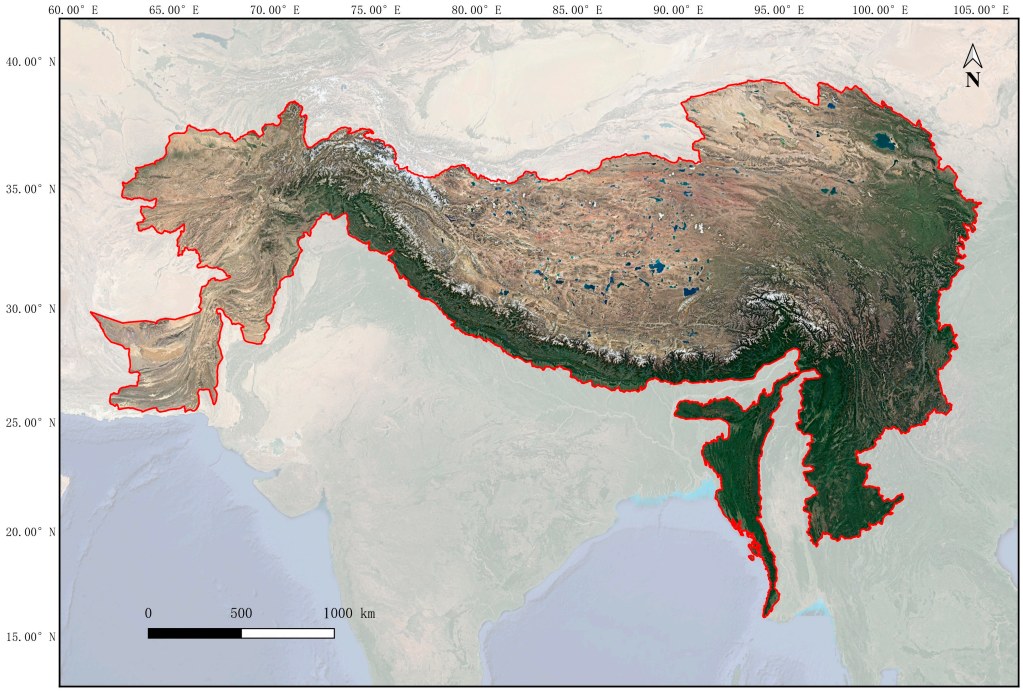

Figure 1. The location of the Hindu Kush Himalaya.

2.2. Data Sources

Sentinel-2 remote sensing imagery, ALOS World 3D terrain data, and the European Space Agency (ESA) WorldCover 10 m 2020 product were the main data sources used in this study. They are described as follows:

(1) Sentinel-2 remote sensing imagery

Sentinel-2 imagery, with its high spatiotemporal resolution, was the primary source used in our study to extract water surface data for the HKH region. It contains 13 spectral bands, with spatial resolutions of 10 m, 20 m, and 60 m, and a 5-day revisiting period. It is one of the most popular data sources for land monitoring [44]. In this study, Short-wave Infra-Red 1 (SWIR1) with 20 m resolution, Near Infra-Red (NIR) with 10 m resolution, and Red with 10 m resolution were selected as the input bands for the water surface extraction network due to their sensitivity to water surfaces [45]. The representativeness of different water surface and non-water objects was comprehensively considered when selecting

images from the Sentinel-2 data source, and only the images with less than 20% cloud cover were considered in order to ensure good quality in the samples generated based on them. The Sentinel-2 images we used were the L2A products of Sentinel-2, and they were Bottom-of-Atmosphere (BOA) reflectance images that underwent pre-processing, including radiometric processing, atmospheric correction, etc.

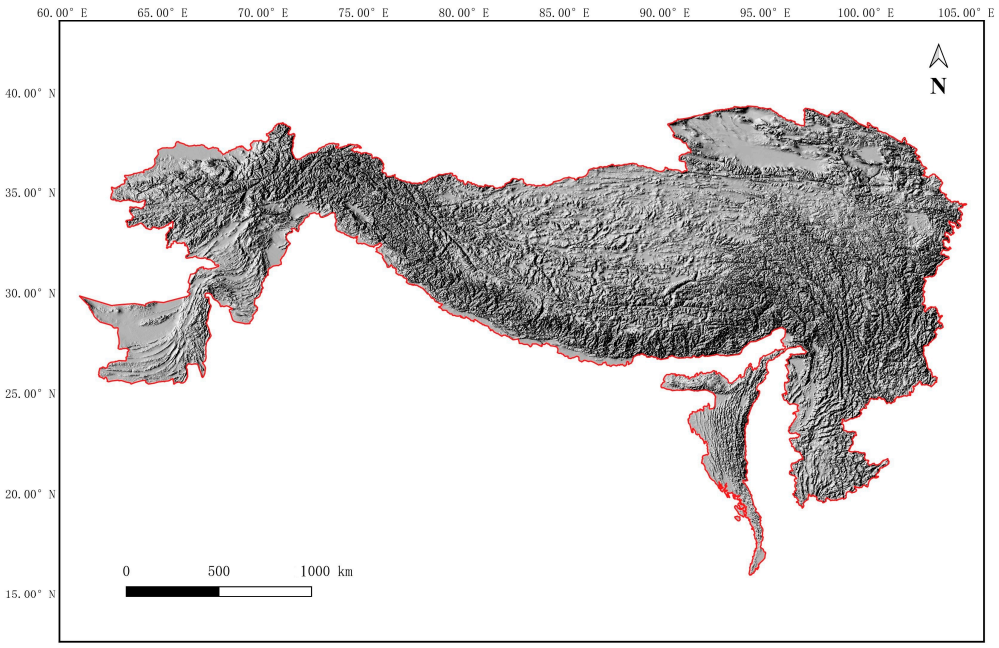

Figure 2. The terrain diagram of the Hindu Kush Himalaya.

(2) ALOS World 3D terrain data

ALOS World 3D-30m (AW3D30) is a digital surface model (DSM) of the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM), which was an optical sensor on board the Advanced Land Observing Satellite "ALOS", provided by the Japan Aerospace Exploration Agency (JAXA). It contains not only the elevation information of the terrain but also includes the height information of surface buildings, trees, and so on [46]. It has a horizontal resolution of 30 m and an elevation accuracy of 5 m. It is one of the most commonly used and accurate terrain data products [47], and we used it as a data source for terrain features in the HKH region. There are several historical versions of this product, and the version we used is version 2.1.

(3) European Space Agency (ESA) World Cover 10 m 2020 product

The ESA World Cover data product was used to accelerate labeling water surfaces in the training dataset. This product is a global land cover product with 10 m resolution [48], which is generated based on Sentinel-1 and Sentinel-2 data. It has high overall accuracy and detailed categories, and it is one of the most popular global land cover products, featuring the following 11 land cover categories: "Tree cover", "Shrubland", "Grassland", "Cropland", "Built-up", "Bare/sparse vegetation", "Snow and Ice", "Permanent water bodies", "Herbaceous Wetland", "Mangrove", and "Moss and lichen". In this study, we extracted the category "Permanent water bodies" to assist in water surface labeling.

3. Methods

3.1. Water Surface Extraction Network

The water surface extraction network proposed in this study uses encoder−decoder architecture [49]. The overall network structure is shown in Figure 3. To extract water surfaces of different types and sizes in the HKH region, the network was built on a Swin Transformer [50], which is based on an attentional mechanism [51]. It imitates the operating mechanism of the human brain when observing things, focuses on important information,

and it can extract water surface features at various levels, from texture to scene [52], allowing it to better capture the differences between water surfaces and non-water objects in the HKH region.

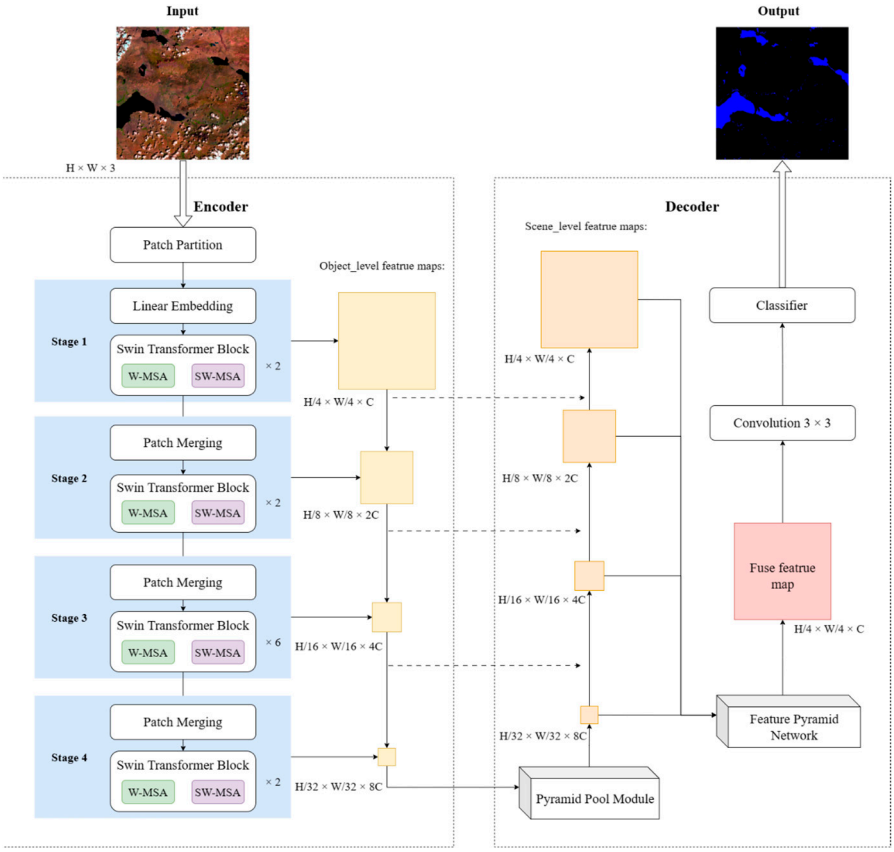

Figure 3. The structure of the water surface extraction network.

The encoder consists of four stages. After the input remote sensing image, with a size of $H \times W \times 3$, is divided into patches through the patch partition module, it is processed through these four stages in sequence. Each stage includes a Swin Transformer Block and a Linear Embedding or Patch Merging Module. The Swin Transformer Block consists of Window Multihead Self-Attention (W-MSA) and Shifted-Window Multihead Self-Attention (SW-MSA). The W-MSA calculates local self-attention within the windows, while the SW-MSA calculates global self-attention by shifting to interact with the information. In this first stage, the image is transformed into a one-dimensional vector via the Linear Embedding Module. In stages 2−4, the Patch Merging Module downscales the feature maps to form pyramid feature maps. Each stage of processing will generate an object-level feature map, so a total of four feature maps can be obtained, with sizes of $H/4 \times W/4 \times C$, $H/8 \times W/8 \times 2C$, $H/16 \times W/16 \times 4C$, and $H/32 \times W/32 \times 8C$.

The decoder consists of a Pyramid Pooling Module (PPM) [53], a Feature Pyramid Network (FPN) [54], a 3×3 convolution, and a classifier. Scene-level feature maps are produced with the PPM and combined with the object-level feature maps from the encoder with the FPN to form a fused feature map. With the 3×3 convolution, the fused feature map is then upsampled into the original image size, and the classifier is applied to perform a pixel-level classification output.

The Transformer-based water surface extraction network has 62.3 million parameters, and the embedding dim is set to 96, the patch size is set to 2, the window size is set to 9×9, and the number of heads for the multihead attention mechanisms in the four sequential stages are set to 3, 6, 12, and 24, respectively.

In addition to the water surface extraction network proposed in this study, we also utilized two CNNs (U-Net [55] and Deeplab V3 with a ResNet-50 backbone [56]) to extract water surfaces, and they were compared to the Transformer-based water surface extraction network. This step was conducted to further verify whether the Transformer-based network outperforms the CNNs in water surface extraction.

3.2. Production of Training Datasets

Figure 4 shows the process of producing the training datasets, including image data preparation and ground truth data preparation.

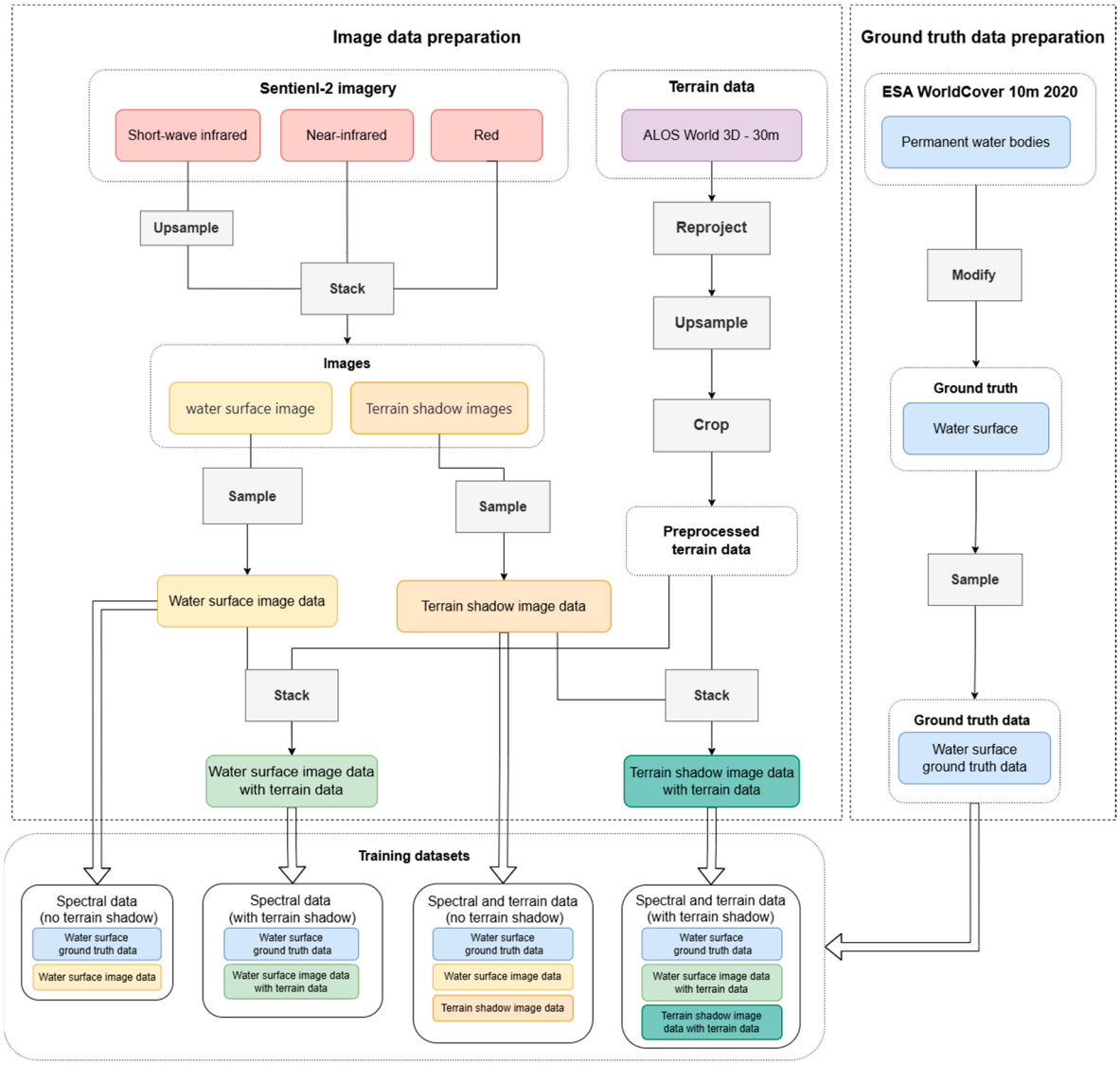

Figure 4. The preparation of training datasets.

During image data preparation, we stacked the SWIR1, NIR, and Red bands of the Sentinel-2 images to obtain the images. To keep the spatial resolution of the model input band consistent, the SWIR1 band was upsampled to 10 m resolution using the nearest neighbor algorithm, and the images were divided into water surface images and terrain shadow images. The water surface images contained mainly water surfaces and sample backgrounds with sparse vegetation, while the terrain shadow images contained different types of terrain shadows; both types of images were manually classified and selected. To ensure that the water surface features in the two training datasets were consistent in water

surface distribution, the selected terrain shadow images did not contain any water. We sampled from these images, generated image data of specified sizes, and obtained both water surface image data and terrain shadow image data. The utilized image size was 768 pixels × 768 pixels.

To investigate whether terrain data have significant impacts on the Transformer-based water surface extraction model, we stacked the spectral images and the pre-processed ASWD3D-30m DSM data as inputs for the model. The specific pre-processing of the DSM data included reprojection, upsampling, and cropping. The 30 m DSM data were reprojected and upsampled to 10 m in order to match the spatial resolution of the 10 m Sentinel-2 data. In addition, the DSM data were cropped based on the corresponding boundaries of the image data. The preprocessed ASWD3D-30m DSM data were stacked with the water surface image data to obtain combined water surface image and terrain data, and they were stacked with the terrain shadow image data to obtain combined terrain shadow image and terrain data.

For ground truth data preparation, we extracted the "Permanent water bodies" category layer from the ESA WorldCover 10 m 2020 product, and then we manually modified the layer based on the selected Sentinel-2 images, which included deleting redundant water surface objects that did not exist in the images, adding water surface objects that existed in the images, and modifying the specific boundaries of the water surfaces; thereby, accurate water surface ground truth images corresponding to each Sentinel-2 image were obtained. The water surface ground truth images were also cropped based on the corresponding image data to produce ground truth data with the same spatial range as the image data.

Finally, we produced the four following training datasets, as shown in Table 1: spectral data (no terrain shadow), spectral data (with terrain shadow), spectral and terrain data (no terrain shadow), and spectral and terrain data (with terrain shadow). In this experiment, the number of samples containing terrain shadows was 352, accounting for 16.8% of the total number of samples in the dataset with terrain shadow samples. It can be calculated from the ground truth data that the number of water surface pixels accounts for approximately 8.6% of the total number of pixels in the dataset without terrain shadow samples and approximately 7.2% of the total number of pixels in the dataset with terrain shadow samples.

Table 1. The four training datasets utilized in the experiment.

Training Dataset	With Terrain Shadow Samples	With Terrain Data	Number of Samples NOT Containing Terrain Shadows	Number of Samples Containing Terrain Shadows	Ratio of Samples Containing Terrain Shadows
Spectral data (no terrain shadow)	No	No	1742	0	0%
Spectral data (with terrain shadow)	Yes	No	1742	352	16.8%
Spectral and terrain data (no terrain shadow)	No	Yes	1742	0	0%
Spectral and terrain data (with terrain shadow)	Yes	Yes	1742	352	16.8%

Using the same method, we produced two validation datasets, one with terrain data and the other without terrain data. The validation datasets with terrain data were designed to evaluate the accuracy of the models trained using the training datasets with terrain data; the validation datasets without terrain data were designed to evaluate the accuracy of the models trained using the training datasets without terrain data. To ensure fairness in the accuracy evaluation, the number and distribution of samples in the two validation datasets were the same, and they both covered different types of water surfaces and terrain shadows in multiple areas.

3.3. Training Settings

The proposed network was implemented using the PyTorch framework (version 1.13.1), and the hardware environment was 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 and

12 GB memory. The optimizer for training was adaptive moment estimation (ADAM) with a total of 300 training epochs. A linear learning rate was applied, starting at 0.000003 and increasing to 0.00006 through linearly changing the small multiplicative factor after 10 epochs. In order to prevent overfitting, we set the weight decay to 0.01 and set the batch size to 4 based on the GPU memory capacity.

Under this hardware environment, using all training samples to train the model for one epoch took about 15 min, and each model was trained for 300 epochs. During the prediction phrase, extracting the water surface result from a Sentinel-2 image took about 90 s.

3.4. Accuracy Evaluation

Overall Accuracy (OA), Intersection over Union (IoU), and Kappa were used as the metrics with which to quantitatively evaluate the accuracy levels of the trained models.

OA and IoU are calculated with the confusion matrix, composed of True-Positives (TP), False-Positives (FP), False-Negatives (FN), and True-Negatives (TN), which, respectively, represent the correctly predicted number of water surface pixels, the incorrectly predicted number of water surface pixels, the unpredicted number of water surface pixels, and the correctly predicted number of non-water surface pixels. The equations for OA and IoU can be represented as follows:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{1}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{2}$$

Kappa is used to measure the consistency between the number of predicted pixels of elements of different classes and the real number of pixels [57]; its equation is as follows:

$$\mathrm{Kappa} = \frac{\mathrm{Po} - \mathrm{Pe}}{1 - \mathrm{Pe}} \tag{3}$$

Po has the same meaning as OA; the equation of Pe is as follows:

$$\mathrm{Pe} = \frac{\mathrm{a1} \times \mathrm{b1} + \mathrm{a2} \times \mathrm{b2} + \cdots + \mathrm{ax} \times \mathrm{bx}}{\mathrm{n} \times \mathrm{n}} \tag{4}$$

In the equation, a1, a2... ax represent the numbers of true pixels in each class; b1, b2..., bx represent the predicted numbers of samples in each class; n is the total number of pixels.

4. Results

4.1. Accuracy Evaluation Results

The evaluation results of the water surface extraction models trained by the training dataset "Spectral data (no terrain shadow)" in Section 3.2 are shown in Table 2. In order to avoid contingency and uncertainty in the results, the scores of the top three accuracy epochs are shown, and their average scores are calculated. "Model_Transformer_Baseline" in Table 2 is a baseline model trained using the samples without terrain data (TD) or with terrain shadow samples (TS). The average OA, IoU, and Kappa scores were 0.9985, 0.9816, and 0.9899, respectively, and they showed that the proposed Transformer-based deep learning network can extract water surfaces from Sentinel-2 images successfully. Additionally, the following two models in Table 2 were obtained by training U-Net and Deeplab V3 using the same training dataset. The average OA, IoU, and Kappa scores for the Model_U-Net_Baseline and Model_Deeplab V3_Baseline were 0.9781, 0.8573, 0.7689 and 0.9964, 0.9563, 0.9757, respectively. This proves that, under the conditions of the experiment conducted in this study, the Transformer-based network has higher accuracy for water surface extraction than the widely used CNN networks.

Table 2. The evaluation results of the model obtained by training the Transformer-based network using the training dataset "Spectral data (no terrain shadow)".

Model	Training Dataset	Epoch	OA	IoU	Kappa
Model_Transformer_Baseline	Spectral data (no terrain shadow)	270 (Top 1)	0.9985	0.9818	0.9900
		202 (Top 2)	0.9985	0.9815	0.9898
		285 (Top 3)	0.9985	0.9815	0.9898
		average	0.9985 *	0.9816 *	0.9899 *
Model_U-Net_Baseline	Spectral data (no terrain shadow)	299 (Top 1)	0.9787	0.8599	0.7724
		288 (Top 2)	0.9779	0.8562	0.7673
		298 (Top 3)	0.9776	0.8559	0.7669
		average	0.9781	0.8573	0.7689
Model_Deeplab V3_Baseline	Spectral data (no terrain shadow)	215 (Top 1)	0.9966	0.9589	0.9772
		144 (Top 2)	0.9964	0.9563	0.9757
		294 (Top 3)	0.9962	0.9537	0.9742
		average	0.9964	0.9563	0.9757

The "*" indicates the average value with the highest accuracy.

Table 3 shows the evaluation results of the four models obtained by training the proposed Transformer-based water surface extraction network using four different training datasets,

Table 3. The evaluation results of the model obtained by training the Transformer-based network using different training datasets.

Model	Training Dataset	Epoch	OA	IoU	Kappa
Model_Transformer_Baseline	Spectral data (no terrain shadow)	270 (Top 1)	0.9985	0.9818	0.9900
		202 (Top 2)	0.9985	0.9815	0.9898
		285 (Top 3)	0.9985	0.9815	0.9898
		average	0.9985 *	0.9816	0.9899
Model_Transformer_TS	Spectral data (with terrain shadow)	195 (Top 1)	0.9986	0.9825	0.9904
		285 (Top 2)	0.9985	0.9822	0.9902
		271 (Top 3)	0.9985	0.9821	0.9902
		average	0.9985 *	0.9823 *	0.9903 *
Model_Transformer_TD	Spectral and terrain data (no terrain shadow)	267 (Top 1)	0.9981	0.9730	0.9853
		206 (Top 2)	0.9980	0.9723	0.9849
		172 (Top 3)	0.9980	0.9722	0.9848
		average	0.9980	0.9725	0.9850
Model_Transformer_TS+TD	Spectral and terrain data (with terrain shadow)	216 (Top 1)	0.9981	0.9734	0.9855
		242 (Top 2)	0.9981	0.9733	0.9854
		146 (Top 3)	0.9981	0.9727	0.9851
		average	0.9981	0.9730	0.9853

The "*" indicates the average value with the highest accuracy.

"Model_Transformer_Baseline" in Table 3 is the same as "Model_Transformer_Baseline" in Table 2.

"Model_Transformer_TS" in Table 3 was trained with the samples containing terrain shadow images. It demonstrated the highest average OA, IoU, and Kappa scores of 0.9985, 0.9823, and 0.9903, respectively. Thus, adding terrain shadow images to the training samples can improve the accuracy of water surface extraction results based on the proposed Transformer-based network.

"Model_Transformer_TD" in Table 3 was trained with the samples containing DSM data. The average OA, IoU, and Kappa scores were 0.9980, 0.9725, and 0.9850, respectively, lower than those of the Model_Transformer_Baseline, thus demonstrating that the introduction of additional terrain data reduces the accuracy of water surface extraction results based on the proposed Transformer-based network.

When terrain data were added to the dataset with terrain shadow samples, the average OA, IoU, and Kappa scores of "Model_Transformer_TS+TD" became 0.9981, 0.9730, and 0.9853, reducing the accuracy of "Model_Transformer_TS". Compared with "Model_Transformer_TD", "Model_Transformer_TS+TD" had higher scores, which also shows that adding terrain shadow samples can improve the accuracy of the model.

4.2. Results of the Misclassification of Terrain Shadows as Water Surfaces

The misclassification of terrain shadows as water surfaces was analyzed on the basis of the two following area scenes: unvegetated areas and vegetated areas.

(1) Misclassification of terrain shadows in unvegetated areas

The misclassification of terrain shadows in unvegetated areas is shown in Figure 5. When "Model_Transformer_Baseline" was used to extract water surfaces, many terrain shadows in the bare area, regardless of their sizes, were misclassified as water surfaces (Figure 5b), thus showing that, when the training dataset does not contain terrain shadow samples and terrain data, the trained model cannot distinguish water surfaces and terrain shadows in the bare area.

When "Model_Transformer_TS" was used to extract water surfaces, almost no terrain shadows in unvegetated areas were misclassified as water surfaces (Figure 5c), thus showing that adding terrain data can improve the ability of the trained model to distinguish terrain shadows and water surfaces in unvegetated areas.

When "Model_Transformer_TD" was used to extract water surfaces, no terrain shadows in the unvegetated areas were misclassified as water surfaces, similar to when "Model_Transformer_TS" was used (Figure 5d), showing that, in unvegetated areas, the addition of terrain data to the training dataset can also reduce the interference of terrain shadows in water surface extraction.

(2) Misclassification of terrain shadows in vegetated areas

The misclassification of terrain shadows as water surfaces in vegetated areas is shown in Figure 6. When "Model_Transformer_Baseline" was used to extract water surfaces, a small number of terrain shadows in areas with high vegetation coverage were misclassified as water surfaces (Figure 6b), fewer than the misclassified terrain shadows in the unvegetated areas.

When "Model_Transformer_TS" was used to extract water surfaces, no terrain shadows in the vegetated areas were misclassified as water surfaces (Figure 6c), thus showing that, whether in unvegetated or vegetated areas, adding terrain shadow samples to the training dataset can improve the model's performance in distinguishing terrain shadows from water surfaces.

When "Model_Transformer_TD" was used to extract water surfaces, many terrain shadows in the vegetated areas were misclassified as water surfaces (Figure 6d), and the areas of misclassified terrain shadows were larger than those resulting from using "Model_Transformer_Baseline", showing that, in vegetated areas, adding terrain data affects the model's performance in distinguishing between terrain shadows and water surface.

4.3. Water Surface Extraction Results

Figure 7 shows the water surface extraction results. We present the results in terms of rivers, small lakes, medium lakes, and large lakes. Small lakes refer to lakes whose area is less than 1 km^2, and the large lakes refer to lakes whose area is greater than 100 km^2; the rest of the lakes are medium lakes.

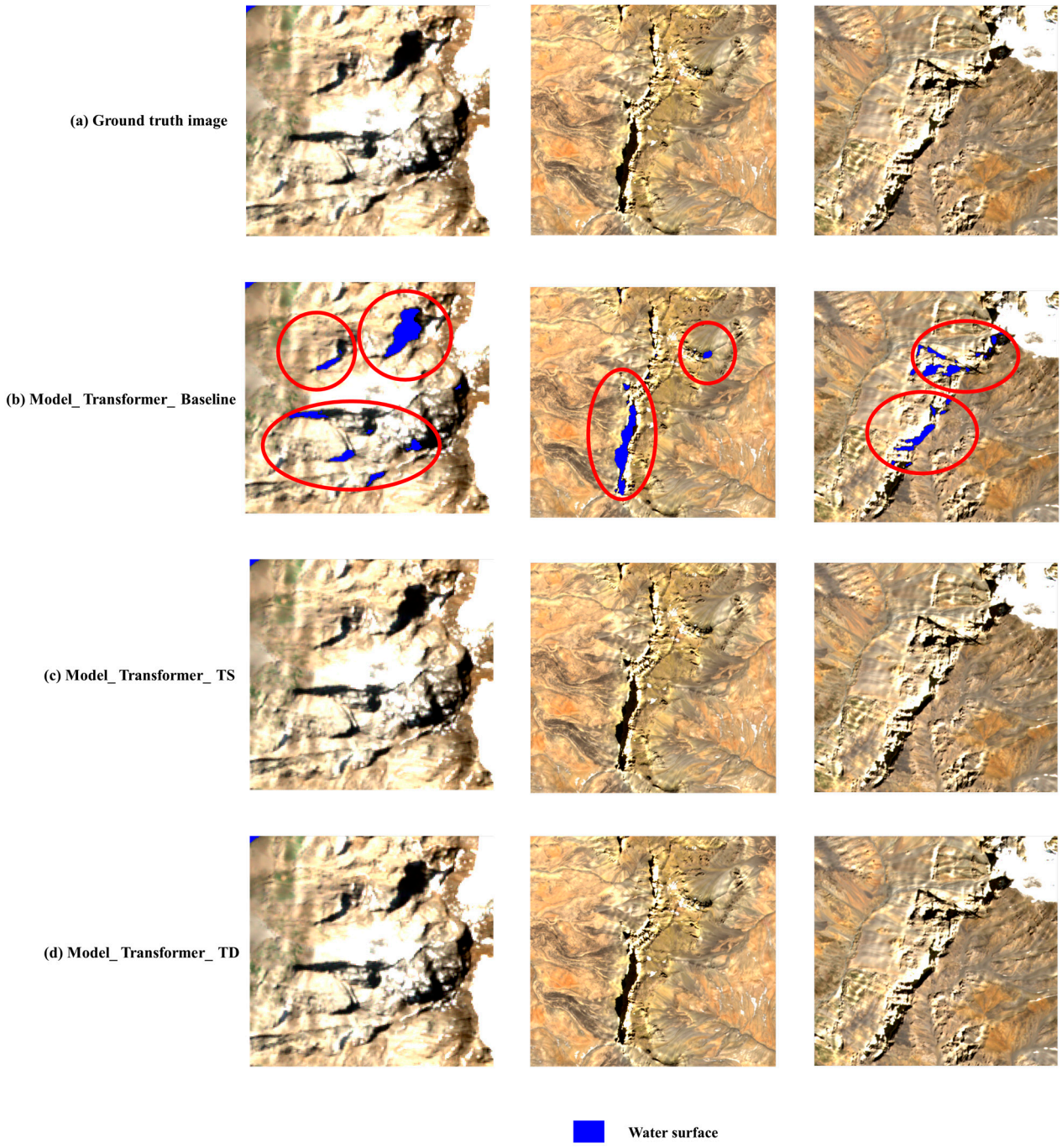

Figure 5. Misclassification of terrain shadows in unvegetated areas (areas circled in red represent terrain shadows misclassified as water surfaces).

When using "Model_Transformer_Baseline" to extract rivers, most of the river pixels were present in the extraction result, but a small number of pixels were missed (Figure 7a). When using "Model_Transformer_TS" to extract rivers, the extracted river morphology was more complete than that extracted with "Model_Transformer_Baseline". However, when "Model_Transformer_TD" was used to extract rivers, the river surface was basically not extracted. Since "Model_Transformer_TD" was obtained by adding the terrain data based on "Model_Transformer_Baseline", which suggests that the use of terrain data significantly affects the accuracy of river extraction. When "Model_Transformer_TS+TD" was used to extract rivers, it could only extract a small number of river pixels. This demonstrates that,

even if there are terrain shadow samples in the training dataset, the use of terrain data still weakens the model's performance in extracting rivers.

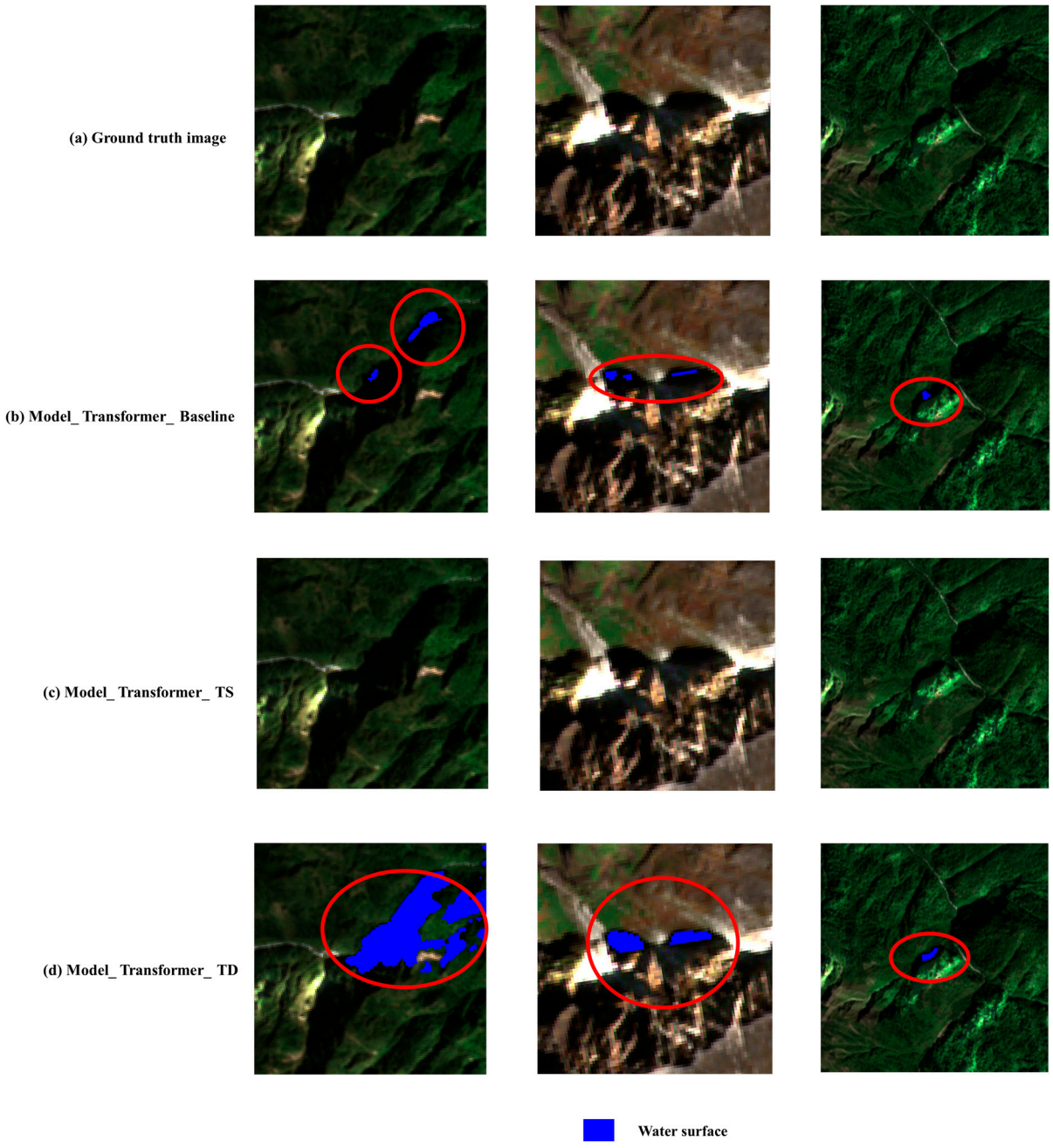

Figure 6. Misclassification of terrain shadows in vegetated areas (areas circled in red represent terrain shadows misclassified as water surfaces).

"Model_Transformer_Baseline" could accurately extract small, medium, and large lakes, with clear boundaries and complete lake surfaces, and no lakes were missed (Figure 7b–d). When "Model_Transformer_TS" was used to extract lakes, the results were the same as those of "Model_Transformer_Baseline", which indicates that adding terrain shadow samples to the training dataset does not affect lake extraction. However, when "Model_Transformer_TD" was used to extract lakes, many small and medium lake pixels were missed, a small number of large lake pixels were missed, and some non-water pixels were extracted, thus showing that adding terrain data affects the accuracy of lake extraction. When "Model_Transformer_TS+TD" was used to extract small lakes, the results were similar to those of "Model_Transformer_TD",

thus proving that, even if the training dataset contains terrain shadow samples, adding terrain data to the training dataset still affects the accuracy of lake extraction.

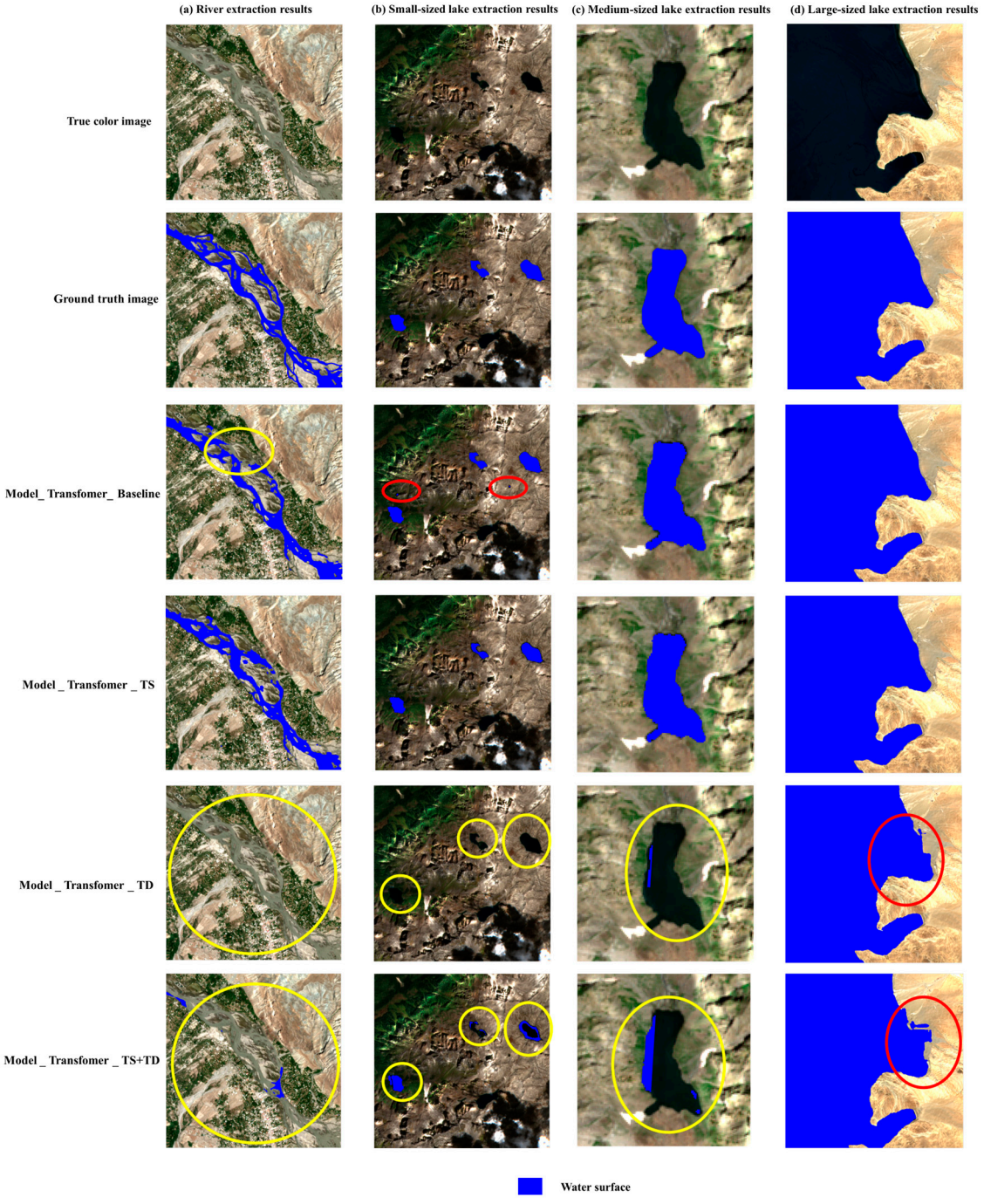

Figure 7. Water surface extraction results (areas circled in yellow represent missed water surface pixels, and areas circled in red represent non-water objects misclassified as water surfaces).

4.4. Results of the Misclassification of Other Non-Water Objects

The results of the misclassification of non-water objects as water surfaces are shown in Figure 8. When using "Model_Transformer_Baseline" and "Model_Transformer_TS" to extract water surfaces, non-water objects, other than terrain shadows, were not misclassified as water surfaces, thus showing that the proposed water surface extraction network could well distinguish water surface from non-water objects and that adding terrain shadow

samples to the training dataset does not affect the model's performance in distinguishing water surfaces from non-water objects.

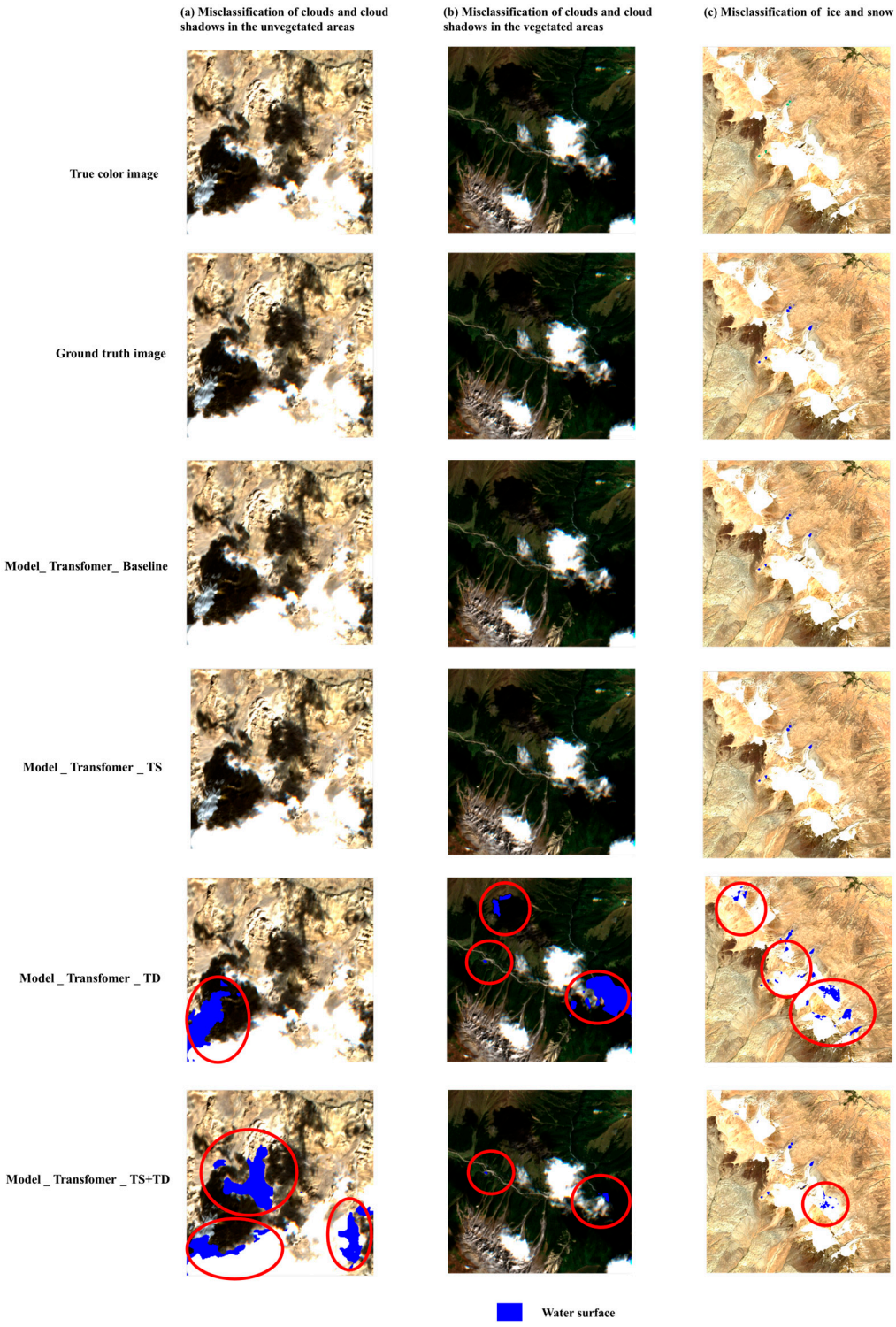

Figure 8. Misclassification of non-water objects (areas circled in red represent non-water objects misclassified as water surface).

However, when using "Model_Transformer_TD" to extract water surfaces, we found that some non-water objects were misclassified as water surfaces, mainly clouds, cloud shadows, ice, and snow, as shown in Figure 8b,c, which shows that adding the terrain data to the training dataset affects the model's performance in distinguishing water surfaces from

non-water objects. In addition, when "Model_Transformer_TS+TD" was used to extract water surfaces, the extraction results were similar to the results of "Model_Transformer_TD", which indicates that, even when the training dataset contains terrain shadow samples, adding terrain data still affects the model's performance in distinguishing water surfaces from non-water objects.

5. Application

5.1. Results of Water Surface Extraction in the HKH Region

In order to validate the robustness of the water surface extraction network, "Model_Transformer_TS", which demonstrated the best performance among the experimental models, was used to extract water surface from Sentinel-2 images of the HKH region; as shown in Figure 9, this region covers 594 tiles of Sentinel-2 images. Since the rainy season in the HKH region is from July to October, we used only the images with less than 20% cloud cover during this period in 2021 to extract water surfaces. The model directly extracted water surfaces from these images without any pre- or post-processing to refine the extraction results. After the extraction, we used a maximum-area algorithm to composite the water surface extraction results with different dates from the same tile and merged the composited results to form the overall water surface extraction results for the HKH region, as shown in Figure 10.

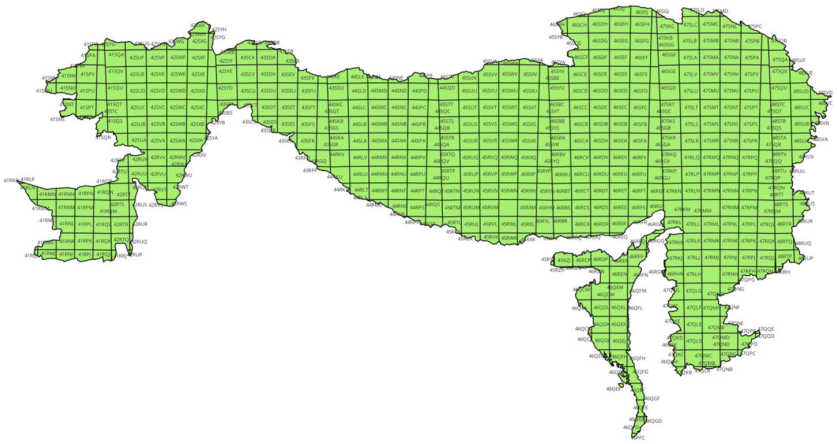

Figure 9. The distribution of Sentinel-2 imaging tiles in the HKH region.

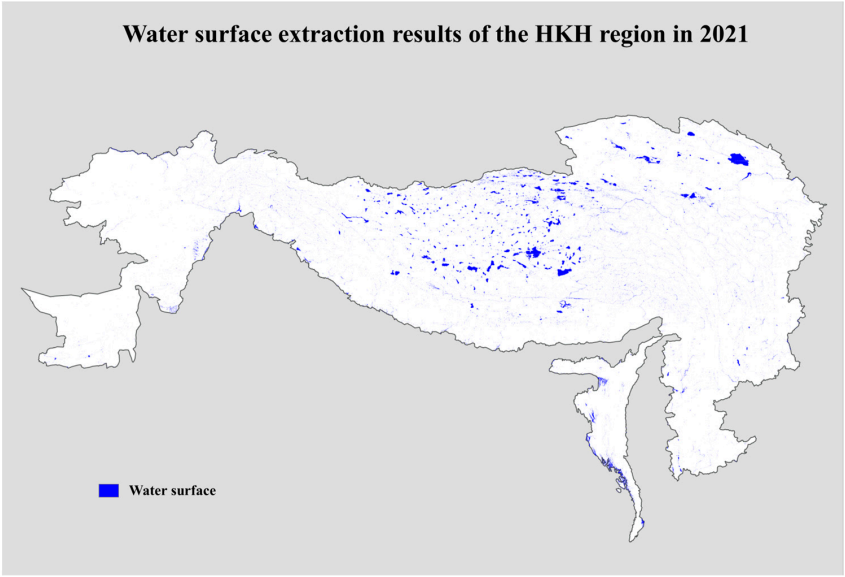

Figure 10. Water surface extraction results of the HKH region in 2021.

5.2. Accuracy Evaluation

To evaluate our water surface results in the HKH region, we compared our water surface extraction results with the Global Surface Water (GSW) seasonality product, one of the most accurate and widely used water surface products, with a resolution of 30 m. The comparison includes a quantitative comparison and an extraction result comparison.

5.2.1. Quantitative Comparison

Table 4 shows the results of the comparison between our results and those of the GSW. It can be seen that the consistency proportions of the selected lakes are all higher than 99%, which shows that our water surface extraction results are highly consistent with the GSW seasonality product on the lakes with large areas and high ecological values.

Table 4. The water surface areas of our extraction results compared with the GSW seasonality product in 2021.

Lake	Area (km^2)		Intersection Area (km^2)	Consistency
	OUR RESULT	GSW Result		
Lake Qinghai	4562.46	4546.04	4544.36	99.60%
Lake Selincuo	2462.65	2452.80	2448.14	99.33%
Lake Namtso	2034.37	2030.60	2024.91	99.54%
Lake Zhari Namco	1055.76	1061.19	1052.83	99.72%
Lake Ngangla Ringco	524.47	530.65	521.55	99.44%

5.2.2. Quantitative Extraction Result Comparison

To further compare our extraction results with those of the GSW seasonality product, we compared them based on the spatial distribution of the water surface using visual interpretation.

Figure 11 shows our water surface extraction results and the GSW seasonality product of the whole HKH region. It can be seen that our extraction results are consistent with the GSW seasonality product in both the spatial distribution of water surface and the whole water surface morphology, with no redundant or abnormal segments in our data, which shows that terrain shadows were not misclassified as water surface in our large-scale extraction results.

In addition, we also compared our extraction results with the GSW seasonality product for some specific water surfaces, including those of large lakes, medium lakes, small lakes, and rivers, the compared results of which are shown in Figures 12, 13, 14 and 15, respectively. For large lakes and medium lakes, our extraction results are basically consistent with the GSW seasonality product regardless of the distribution, overall shape, or boundaries of the lakes, and they are very clear and accurate. For small lakes and rivers, our extraction results are consistent with the GSW seasonality product in regard to overall shape. Since we used 10 m Sentinel-2 images to extract water surfaces, our extraction results are finer at the boundaries and can better reflect the specific shapes of small lakes and rivers, as shown in Figures 14 and 15.

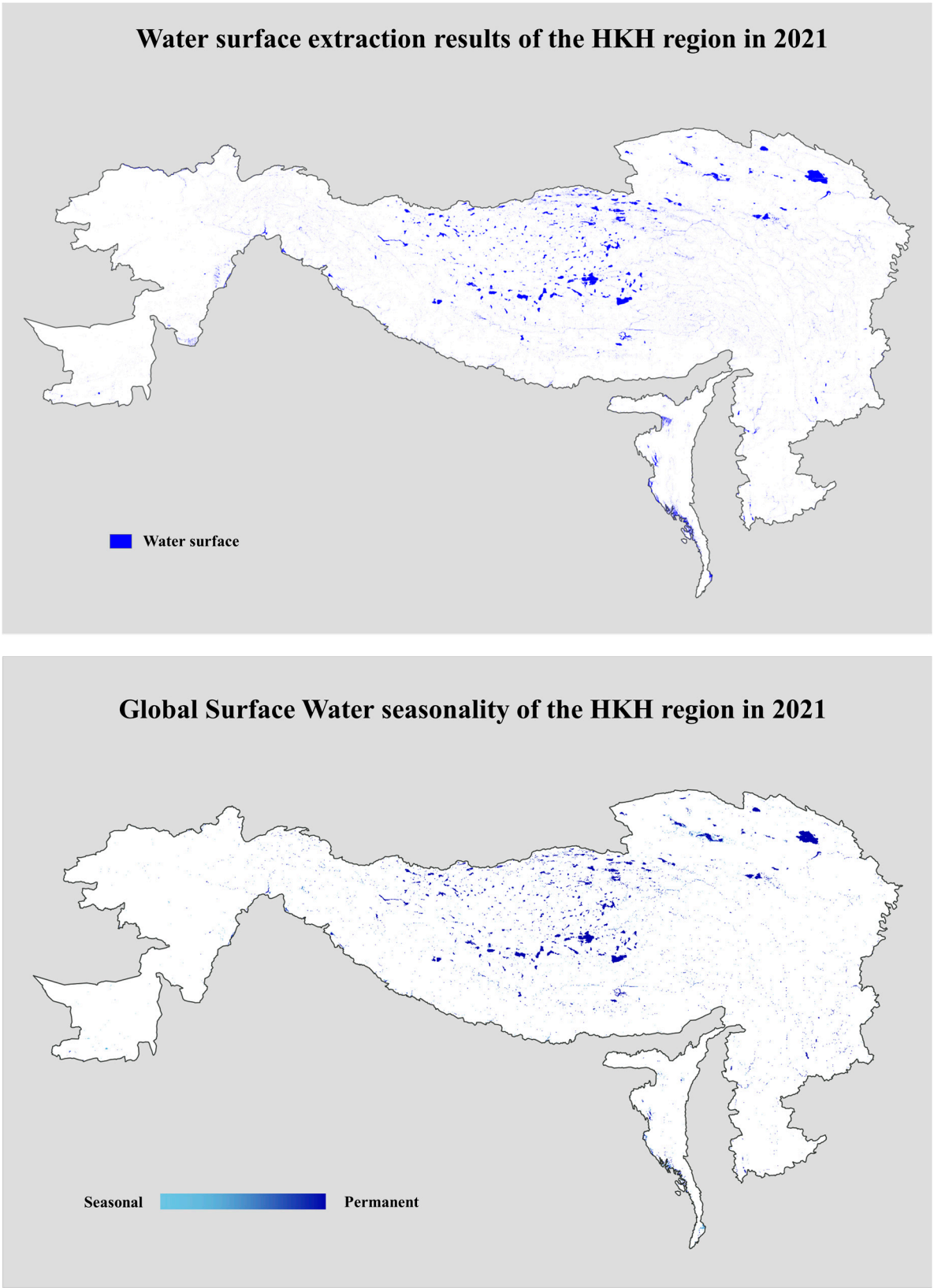

Figure 11. Our water surface extraction results and the GSW seasonality product for the HKH region in 2021.

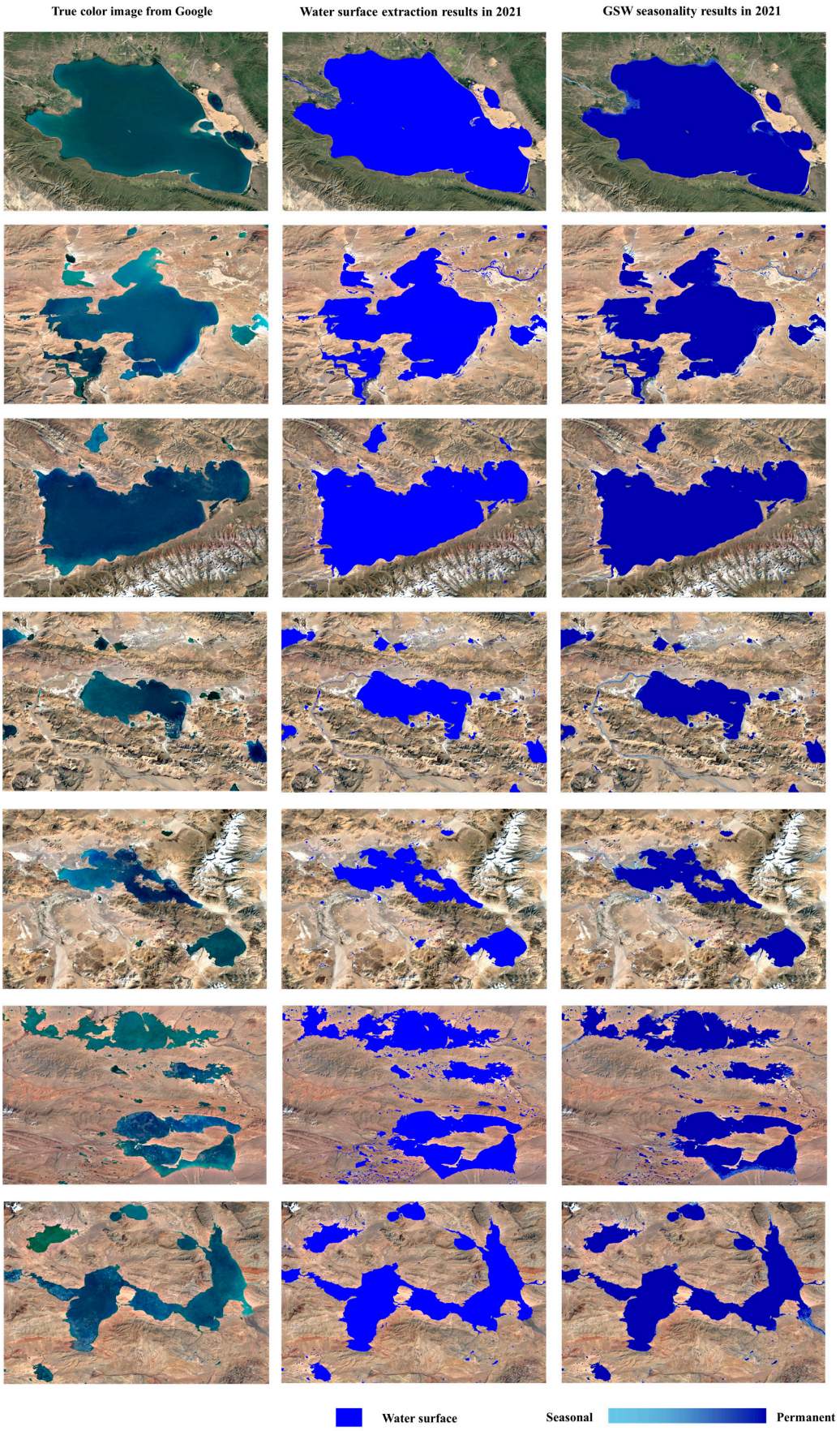

Figure 12. Our water surface extraction results and the GSW seasonality product for large lakes.

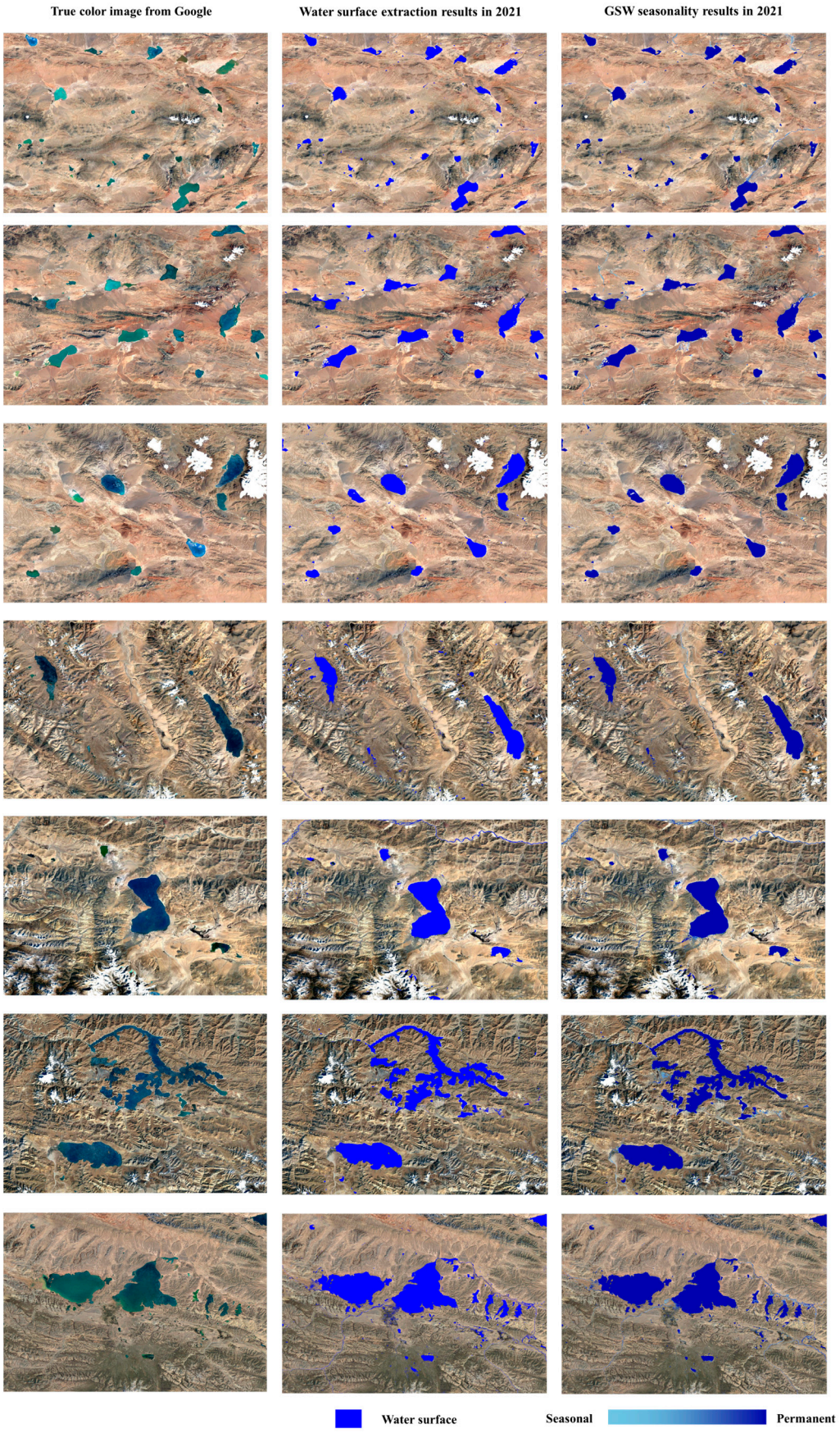

Figure 13. Our water surface extraction results and the GSW seasonality product for medium lakes.

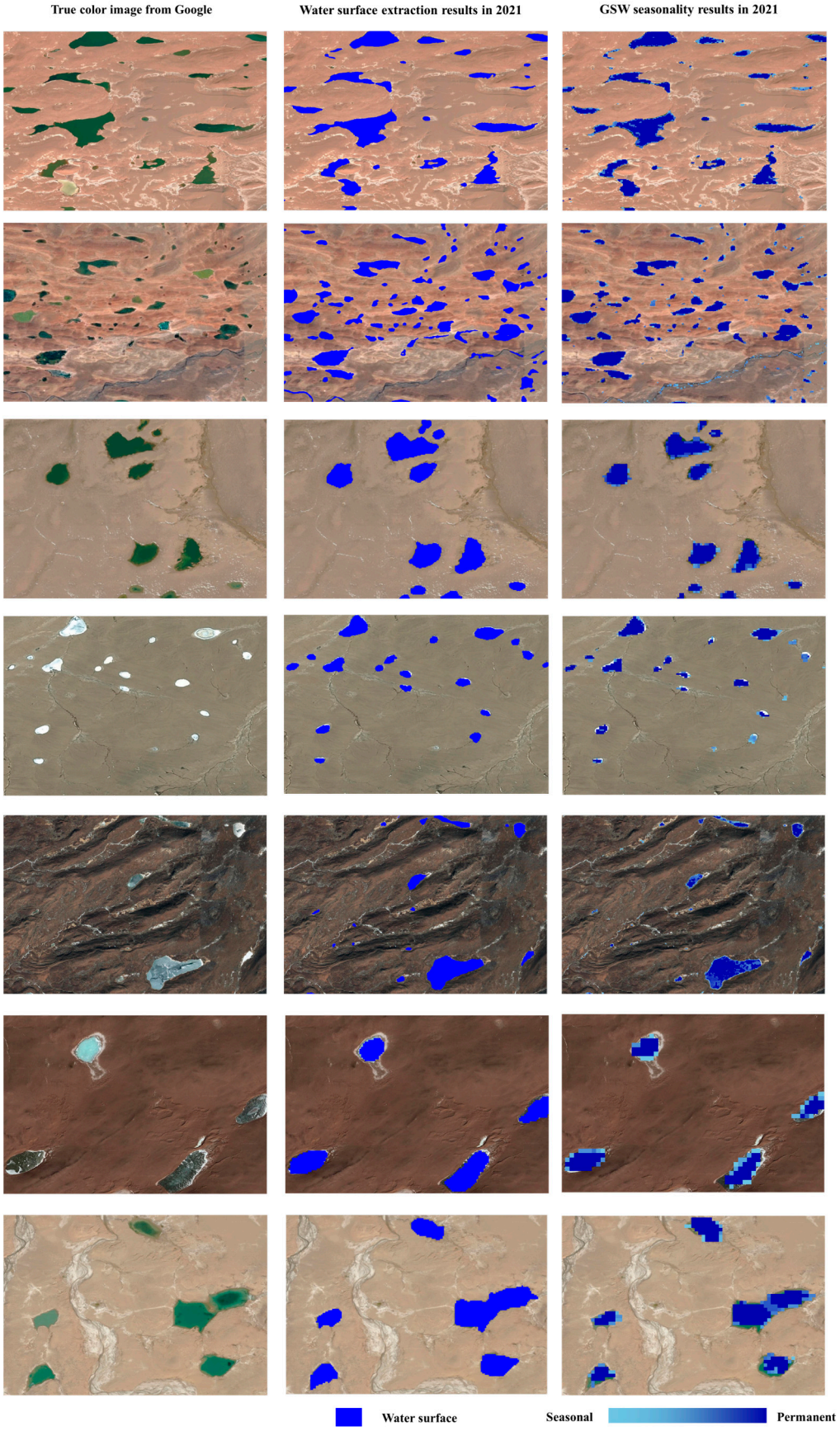

Figure 14. Our water surface extraction results and the GSW seasonality product for small lakes.

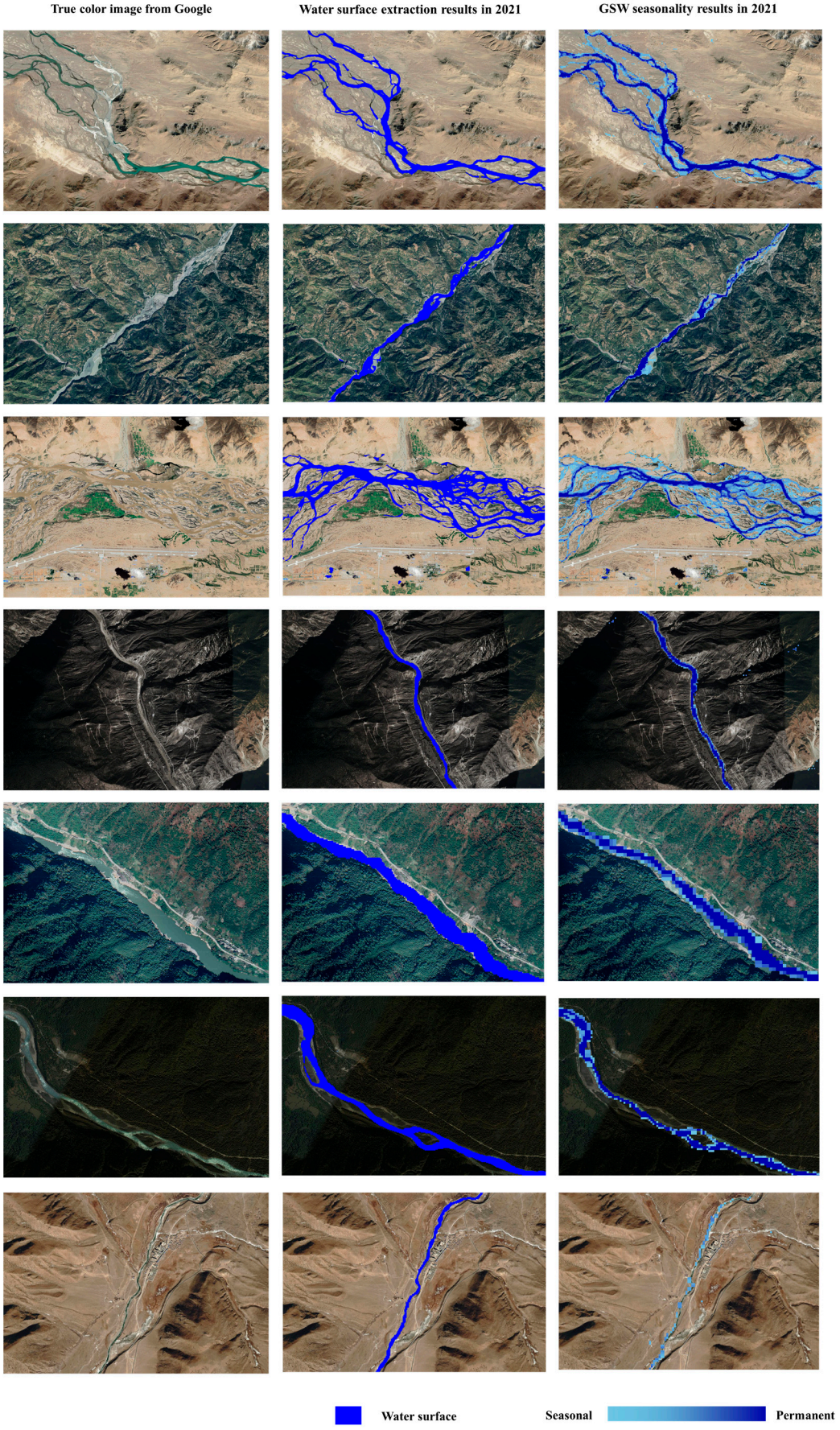

Figure 15. Our water surface extraction results and the GSW seasonality product for rivers.

6. Discussion

In high mountainous areas, such as the HKH region, terrain shadows are an important factor that interferes with water surface extraction; thus, we designed a water surface extraction network based on the Vision Transformer, adding terrain shadow samples to increase the sample representativeness, thereby improving the performance of the network used to distinguish water surface and terrain shadows.

Our experiment shows that, when terrain shadow samples were added to the training dataset, the model's performance in distinguishing water surface and terrain shadows significantly improved and the water surface extraction result became more accurate. This clearly demonstrates that preparing samples containing terrain shadows is a very effective tool when using the Transformer-based deep learning network to reduce the interference of terrain shadows in water surface extraction. Moreover, the proportion of terrain shadow samples in our experiment was low, at about 16.8%. This also confirms that samples containing terrain shadows are extremely important for water surface extraction in high mountainous areas, since the misclassification of terrain shadows as water surfaces can be mostly avoided without adding too many terrain shadow samples. In addition, preparing samples containing terrain shadows (and not having to label them) is much easier than other methods, including adding and processing terrain data. The application case whereby the entire water surface distribution of the HKH region in 2021 was extracted, based on the proposed deep learning method, further strongly validates the effectiveness and usability of Transformer-based deep learning networks with samples containing both water surface and terrain shadows. In addition to terrain shadows, some other non-water objects, such as clouds, cloud shadows, glaciers, and snow, are also potentially misclassified as water surfaces. We believe that these misclassifications can also be reduced by adding a certain number of relevant samples to the training dataset.

In order to investigate whether terrain data significantly help to reduce the interference of terrain shadows during water surface extraction in high mountainous areas, we introduced AW3D30 terrain data and prepared samples that contained terrain data to train the same Transformer-based deep learning network. Our experimental results show that adding the AW3D30 terrain data to the training dataset is not very effective; while it can reduce the misclassification of terrain shadows as water surface in some unvegetated areas, misclassification is not reduced, or becomes even more severe, in vegetated areas. In addition, the performance of the water surface extraction model decreases significantly after adding the terrain data, as demonstrated in the experimental results showing incomplete or incorrect water surface boundaries. We wondered whether the accuracy of the terrain data might have caused the undesirable results above; in reality, however, the terrain data we used were already among those of the highest accuracy publicly available today. Moreover, the economic and labor costs of obtaining more accurate terrain data would be higher. In addition, it is possible that the deep learning networks themselves may perform poorly when faced with training data containing large differences, since terrain data and spectral imagery have two completely different physical features.

Finally, it is important to note that all of the experimental results and conclusions presented in this research are based on the approach of deep learning networks, especially Transformer-based networks; therefore, the method of reducing terrain shadow interference by adding terrain shadow samples may only be effective for water surface extraction using these deep learning networks. It is unclear whether the approach used in our research is applicable to other extraction methods, such as water index or shallow machine learning methods.

7. Conclusions

In this study, we designed a water surface extraction network, based on the Vision Transformer, in order to efficiently and automatically extract water surfaces in high mountainous areas, such as the HKH region, and explored utilizing the method to reduce the interference of terrain shadows during water surface extraction. Our results show that

adding terrain shadow samples can greatly improve the model's performance in distinguishing between water surface and terrain shadows, and the model can accurately extract water surfaces; meanwhile, adding specific terrain data is not very effective in reducing the misclassification of terrain shadows as water surfaces, and this approach could reduce the accuracy of water surface extraction in the HKH region. Using the Transformer-based network and sufficient samples of both water surface and terrain shadows, we quickly obtained water surface extraction results from the HKH region in the year 2021. Comparison of our extraction results with the GSW seasonality product shows that the two are highly consistent, and our extraction results are finer at the water surface boundaries. This sufficiently demonstrates the high spatiotemporal generalization ability of our water surface extraction network, as well as the broad feasibility of reducing terrain shadow interference by simply adding terrain shadow samples. In the future, we plan to produce more samples from different geographical areas and apply the proposed method to various terrain environments beyond the HKH region in order to verify the generalization capability of the method.

Author Contributions: X.Y.: Writing—original draft, Methodology, Validation, Visualization, Data Curation. J.S.: Writing—review and editing, Conceptualization, Methodology, Investigation, Funding Acquisition, Software, Supervision, Resources. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (grant number: 2022YFF0711602, 2021YFE0117800), and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number: XDB0740200).

Data Availability Statement: The water surface distribution of the Hindu Kush Himalaya in 2021 using the proposed method in this study is available at https://zenodo.org/records/10691444, accessed on 22 February 2024. The study obtained the Sentinel-2 images from the Copernicus Open Access Hub (https://scihub.copernicus.eu/, accessed on 12 May 2023), and ALOS World 3D-30m (AW3D30) from the official website of ALOS Research and Application project (https://www.eorc.jaxa.jp/ALOS/en/index_e.htm, accessed on 26 July 2023), and ESA WorldCover 10 m 2020 product from its official website (https://esa-worldcover.org/en/, accessed on 23 May 2023), and the Global Surface Water dataset from the Global Surface Water Explorer (https://global-surface-water.appspot.com/, accessed on 20 September 2023).

Acknowledgments: We appreciate the detailed comments from the editor and the anonymous reviewers, as well as the National Data Sharing Infrastructure of Earth System Science (http://www.geodata.cn/, accessed on 17 May 2023).

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

References

1. Adrian, R.; O'Reilly, C.M.; Zagarese, H.; Baines, S.B.; Hessen, D.O.; Keller, W.; Livingstone, D.M.; Sommaruga, R.; Straile, D.; Van Donk, E.; et al. Lakes as sentinels of climate change. *Limnol. Oceanogr.* **2009**, *54*, 2283–2297. [CrossRef] [PubMed]
2. Moser, K.A.; Baron, J.S.; Brahney, J.; Oleksy, I.A.; Saros, J.E.; Hundey, E.J.; Sadro, S.; Kopáček, J.; Sommaruga, R.; Kainz, M.J.; et al. Mountain lakes: Eyes on global environmental change. *Glob. Planet. Chang.* **2019**, *178*, 77–95. [CrossRef]
3. Molden, D.J.; Vaidya, R.A.; Shrestha, A.B.; Rasul, G.; Shrestha, M.S. Water infrastructure for the Hindu Kush Himalayas. *Int. J. Water Resour. Dev.* **2014**, *30*, 60–77. [CrossRef]
4. Singh, S.; Hassan, S.M.T.; Hassan, M.; Bharti, N. Urbanisation and water insecurity in the Hindu Kush Himalaya: Insights from Bangladesh, India, Nepal and Pakistan. *Water Policy* **2020**, *22*, 9–32. [CrossRef]
5. Wahid, S.M.; Shrestha, A.B.; Murthy, M.S.R.; Matin, M.; Zhang, J.; Siddiqui, O. Regional Water Security in the Hindu Kush Himalayan Region: Role of Geospatial Science and Tools. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*, 1331; -8. [CrossRef]
6. Yang, X.; Lu, X.; Park, E.; Tarolli, P. Impacts of Climate Change on Lake Fluctuations in the Hindu Kush-Himalaya-Tibetan Plateau. *Remote Sens.* **2019**, *11*, 1082. [CrossRef]
7. Li, J.; Warner, T.A.; Wang, Y.; Bai, J.; Bao, A. Mapping glacial lakes partially obscured by terrain shadows for time series and regional mapping applications. *Int. J. Remote Sens.* **2019**, *40*, 615–641. [CrossRef]
8. 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]

9. Qiao, C.; Luo, J.; Sheng, Y.; Shen, Z.; Zhu, Z.; Ming, D. An Adaptive Water Extraction Method from Remote Sensing Image Based on NDWI. *J. Indian Soc. Remote Sens.* **2012**, *40*, 421–433. [CrossRef]
10. 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]
11. Li, C.; Wang, S.; Bai, X.; Tan, Q.; Yang, Y.; Li, Q.; Wu, L.; Xiao, J.; Qian, Q.; Chen, F.; et al. New automated method for extracting river information using optimized spectral threshold water index. *Arab. J. Geosci.* **2019**, *12*, 13. [CrossRef]
12. Li, J.; Meng, Y.; Li, Y.; Cui, Q.; 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]
13. Zhu, X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Trans. Geosci. Remote Sens.* **2017**, *5*, 8–36. [CrossRef]
14. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef] [PubMed]
15. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; Guan, L.; Fang, H. Deep-LearningBased Multispectral Satellite Image Segmentation for Water Body Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7422–7434. [CrossRef]
16. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens.* **2016**, *4*, 22–40. [CrossRef]
17. Chen, Y.; Fan, R.; Yang, X.; Wang, J.; Latif, A. Extraction of Urban Water Bodies from High-Resolution Remote-Sensing Imagery Using Deep Learning. *Water* **2018**, *10*, 585. [CrossRef]
18. 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]
19. 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]
20. Ghosh, S.; Das, N.; Das, I.; Maulik, U. Understanding Deep Learning Techniques for Image Segmentation. *ACM Comput.* **2020**, *52*, 1–35. [CrossRef]
21. Kansizoglou, I.; Bampis, L.; Gasteratos, A. Deep feature space: A geometrical perspective. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 6823–6838. [CrossRef] [PubMed]
22. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523–3542. [CrossRef] [PubMed]
23. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [CrossRef] [PubMed]
24. Horkaew, P.; Puttinaovarat, S. Entropy-Based Fusion of Water Indices and DSM Derivatives for Automatic Water Surfaces Extraction and Flood Monitoring. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 301. [CrossRef]
25. Puttinaovarat, S.; Khaimook, K.; Polnigongit, W.; Horkaew, P. Robust water surface extraction from landsat imagery by using gradual assignment of water index and DSM. In Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 17–19 September 2015; pp. 122–126. [CrossRef]
26. Al-Najjar, H.A.H.; Kalantar, B.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Mansor, S. Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1461. [CrossRef]
27. Wang, X.; Zhou, G.; Lv, X.; Zhou, L.; Hu, M.; He, X.; Tian, Z. Comparison of Lake Extraction and Classification Methods for the Tibetan Plateau Based on Topographic-Spectral Information. *Remote Sens.* **2023**, *15*, 267. [CrossRef]
28. Jiang, X.; Liang, S.; He, X.; Ziegler, A.D.; Lin, P.; Pan, M.; Wang, D.; Zou, J.; Hao, D.; Mao, G.; et al. Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 36–50. [CrossRef]
29. Wu, X.; Zhang, Z.; Xiong, S.; Zhang, W.; Tang, J.; Li, Z.; An, B.; Li, R. A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR Images. *Remote Sens.* **2023**, *15*, 2046. [CrossRef]
30. Yadav, R.; Nascetti, A.; Ban, Y. Deep attentive fusion network for flood detection on uni-temporal Sentinel1 data. *Front. Remote Sens.* **2022**, *3*, 1060144. [CrossRef]
31. Xu, H. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]
32. Yan, X.; Song, J.; Liu, Y.; Lu, S.; Xu, Y.; Ma, C.; Zhu, Y. A Transformer-based method to reduce cloud shadow interference in automatic lake water surface extraction from Sentinel-2 imagery. *J. Hydrol.* **2023**, *620*, 129561. [CrossRef]
33. Song, J.; Yan, X. The Effect of Negative Samples on the Accuracy of Water surface extraction Using Deep Learning Networks. *Remote Sens.* **2023**, *15*, 514. [CrossRef]
34. Aleissaee, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.-S.; Khan, F.S. Transformers in Remote Sensing: A Survey. *Remote Sens.* **2023**, *15*, 1860. [CrossRef]
35. Bazi, Y.; Bashmal, L.; Rahhal, M.M.A.; Dayil, R.A.; Ajlan, N.A. Vision Transformers for Remote Sensing Image Classification. *Remote Sens.* **2021**, *13*, 516. [CrossRef]
36. Chen, K.; Zou, Z.; Shi, Z. Building Extraction from Remote Sensing Images with Sparse Token Transformers. *Remote Sens.* **2021**, *13*, 4441. [CrossRef]

37. Zhang, J.; Zhao, H.; Li, J. TRS: Transformers for Remote Sensing Scene Classification. *Remote Sens.* **2021**, *13*, 4143. [CrossRef]
38. Song, C.; Huang, B.; Ke, L.; Keith, S.R. 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]
39. Shea, J.; Immerzeel, W. An assessment of basin-scale glaciological and hydrological sensitivities in the Hindu Kush–Himalaya. *Ann. Glaciol.* **2016**, *57*, 308–318. [CrossRef]
40. Singh, V.; Pandey, A. Urban water resilience in Hindu Kush Himalaya: Issues, challenges and way forward. *Water Policy* **2020**, *22*, 33–45. [CrossRef]
41. Mukherji, A.; Sinisalo, A.; Nüsser, M.; Garrard, R.; Eriksson, M. Contributions of the cryosphere to mountain communities in the Hindu Kush Himalaya: A review. *Reg. Environ. Chang.* **2019**, *19*, 1311–1326. [CrossRef]
42. You, Q.; Ren, G.; Zhang, Y.; Ren, Y.; Sun, X.; Zhan, Y.; Shrestha, A.; Krishnan, R. An overview of studies of observed climate change in the Hindu Kush Himalayan (HKH) region. *Adv. Clim. Chang. Res.* **2017**, *8*, 141–147. [CrossRef]
43. Chen, J.; Chen, S.; Fu, R.; Li, D.; Jiang, H.; Wang, C.; Peng, Y.; Jia, K.; Hicks, B.J. Remote sensing big data for water environment monitoring: Current status, challenges, and future prospects. *Earth's Future* **2022**, *10*, e2021EF002289. [CrossRef]
44. 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]
45. 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]
46. Tadono, T.; Nagai, H.; Ishida, H.; Oda, F.; Naito, S.; Minakawa, K.; Iwamoto, H. Generation Of The 30 M-Mesh Global Digital Surface Model By Alos Prism. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2016**, *XLI-B4*, 157–162. [CrossRef]
47. Grohmann, C.H. Evaluation of TanDEM-X DEMs on selected Brazilian sites: Comparison with SRTM, ASTER GDEM and ALOS AW3D30. *Remote Sens. Environ.* **2018**, *212*, 121–133. [CrossRef]
48. Zanaga, D.; Van, D.K.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. Ji, Y.; Zhang, H.; Zhang, Z.; Liu, M. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. *Inf. Sci.* **2021**, *546*, 835–857. [CrossRef]
50. 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, BC, Canada, 11–17 October 2021; pp. 9992–10002. [CrossRef]
51. 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]
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. Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; Sun, J. Unified Perceptual Parsing for Scene Understanding. In Proceedings of the Computer Vision–ECCV 2018, Munich, Germany, 8–14 September 2018; Lecture Notes in Computer Science. Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 432–448. [CrossRef]
54. 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]
55. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
56. Chen, L.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. [CrossRef]
57. Cohen, D.; Jordan, S.M.; Croft, W.B. Learning a Better Negative Sampling Policy with Deep Neural Networks for Search. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, Santa Clara, CA, USA, 2–5 October 2019; pp. 19–26. [CrossRef]