

Article

Achieving Higher Resolution Lake Area from Remote Sensing Images Through an Unsupervised Deep Learning Super-Resolution Method

Mengjiao Qin ^{1,2}, Linshu Hu ¹, Zhenhong Du ^{1,3,*}, Yi Gao ¹, Lianjie Qin ^{4,5,6}, Feng Zhang ^{1,3} 
and Renyi Liu ^{1,3}

¹ School of Earth Sciences, Zhejiang University, Hangzhou 310027, China; qinmengjiao@zju.edu.cn (M.Q.); hulinsu1010@zju.edu.cn (L.H.); yi_gao@zju.edu.cn (Y.G.); zfcarnation@zju.edu.cn (F.Z.); liurenyi@zju.edu.cn (R.L.)

² Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, Zhejiang University, Hangzhou 310027, China

³ Key Laboratory of Geographic Information Science of Zhejiang Province, Zhejiang University, Hangzhou 310028, China

⁴ Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; qinlianjie@mail.bnu.edu.cn

⁵ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

⁶ Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management and Ministry of Education, Beijing Normal University, Beijing 100875, China

* Correspondence: duzhenhong@zju.edu.cn

Received: 30 April 2020; Accepted: 12 June 2020; Published: 15 June 2020



Abstract: Lakes have been identified as an important indicator of climate change and a finer lake area can better reflect the changes. In this paper, we propose an effective unsupervised deep gradient network (UDGN) to generate a higher resolution lake area from remote sensing images. By exploiting the power of deep learning, UDG models the internal recurrence of information inside the single image and its corresponding gradient map to generate images with higher spatial resolution. The gradient map is derived from the input image to provide important geographical information. Since the training samples are only extracted from the input image, UDG can adapt to different settings per image. Based on the superior adaptability of the UDG model, two strategies are proposed for super-resolution (SR) mapping of lakes from multispectral remote sensing images. Finally, Landsat 8 and MODIS (moderate-resolution imaging spectroradiometer) images from two study areas on the Tibetan Plateau in China were used to evaluate the performance of UDG. Compared with four unsupervised SR methods, UDG obtained the best SR results as well as lake extraction results in terms of both quantitative and visual aspects. The experiments prove that our approach provides a promising way to break through the limitations of median-low resolution remote sensing images in lake change monitoring, and ultimately support finer lake applications.

Keywords: unsupervised super-resolution; lake; remote sensing; residual network; gradient map

1. Introduction

Lakes are dynamic systems that support enormous biodiversity and provide key provisioning and cultural ecosystem services to people around the world [1]. Since the changes in lakes, such as expansion and shrinkage are closely related to the effect of climate and human activities [2], lakes can act as a salient indicator of environmental change. In recent decades, accelerated climate warming and

rapid economic development have brought about great influence on global lakes. The remote sensing (RS) technique makes long-term and wide-coverage lake monitoring possible. It has been applied to long-term lake evolution [3], lake water storage changes [4], lake level changes [5], etc. However, most studies focused on lakes larger than 10 km² [6–8] due to the limitation of the spatial resolution of RS images. Zhang et al. (2015) indicated that small lakes are more sensitive to climate change because the area changes of small lakes caused by climate change are more significant. As such, generating finer lakes with higher spatial resolution from remote sensing images is of great significance for climate change research.

The super-resolution (SR) technique aims to reconstruct a higher resolution image from its original, low-resolution version, and it has been successfully used in various fields, such as wetland inundation mapping [9], high-resolution digital elevation model (DEM) generation [10,11], remote sensing [12–15] and computer vision [16–19]. Utilizing the SR technique to improve the spatial resolution of the lake area is a promising method, which has advantages of low cost, easy implementation, and high efficiency compared to updating image acquisition devices [13].

The existing SR methods can be roughly divided into two categories: supervised SR and unsupervised SR. The former requires large amounts of low-resolution (LR) images and corresponding high-resolution (HR) images for training [20]. While collecting images of the same scene in high resolutions is very difficult and the image pre-processing and fusion are time-consuming. In addition, the performance of supervised SR methods largely depends on the training samples. Once the test data has different distribution with the training samples, the performance of these models deteriorates significantly [21]. In the contrast, unsupervised SR methods require no matched LR-HR image pairs, which are more flexible to handle different image settings and more likely to cope with the SR problems in real-world scenarios such as generating a higher resolution lake area.

Traditional unsupervised SR methods including bicubic interpolation (BCI), gradient profile prior (GPP) [22], iterative back projection (IBP) [23], and transformed self-exemplars (TSR) [24]. With the development of deep learning, many advanced models, such as deep generative networks [25,26], cycle-in-cycle SR network [27], and “zero-shot” super-resolution (ZSSR) model [21], have been proposed and greatly improved the unsupervised SR performance. ZSSR exploits the internal statistical law within each input image. It trained a small image-specific convolutional neural networks (CNN) at test time, on examples extracted solely from the input image itself. Therefore, ZSSR can adapt to different image settings such as different image channels and image sizes, which can well support the extraction of high-resolution lakes from different multispectral RS images.

However, there are some problems to apply ZSSR in our task directly: (a) the geographic information in RS images such as terrain, structure, edge, has a great impact on lake mapping. Including these information details in the super-resolved HR images are significant; (b) the original RS data may be difficult to collect, sometimes we need to generate higher resolution lakes from products such as the normalized difference water index (NDWI) map provided in public.

Considering all the above, the unsupervised deep gradient network (UDGN) is proposed to generate a higher resolution lake area. The UDGNN model exploits the power of deep learning and consists of feature fusion, deep feature extraction, upsampling and reconstruction modules. The gradient map of the input image is obtained and fused with the input image to provide more geographic details for SR. UDGNN inherits the advantages of ZSSR, i.e., it can handle images with different channels, different sizes. Based on UDGNN, two strategies are designed to flexibly generate lakes with higher resolution from original RS images or intermediate products. Fine lakes mapping with higher resolution can better reflect the effect of climate and human activities. To summarize, the main contributions of this paper are as follows.

- (1) The deep learning-based SR technique is first introduced to the lake area extraction process to improve the spatial resolution and generate a finer lake area.

- (2) A new unsupervised SR model UDGNet is proposed based on a deep residual network in this paper. It does not require pretraining and can be adapted to different settings of images, such as different image sizes and channels.
- (3) The features of the gradient map are extracted and fused in the network to provide more geographic details in HR images.
- (4) We verify the effectiveness of our method with two data sets, the results demonstrate the superiority of our method in improving the spatial resolution of lake area extraction.

2. Materials and Methods

2.1. Study Area and Data

Two study areas are selected from the Tibetan Plateau (TP), as shown in Figure 1. The TP is the largest and highest plateau in the world, and there are numerous lakes distributed throughout the TP [4]. Along with the Arctic region and Antarctica, the Tibetan Plateau (TP) and the Mongolian Plateau (MP) are among the world's most sensitive regions to climate change [28]. Hence, monitoring accurately the changes of lakes in the TP is of great importance for climate change research. In this work, two typical areas located in the TP, China are chosen to evaluate the effectiveness and practicability of our method. Study area 1 has a large tectonic lake (Yamzho Yumco) with irregular edges, which is used to verify the superiority of our proposed SR method when dealing with lakes with complex terrain and intricate structures. In terms of study area 2, there are many small lakes such as Sangzhen and other no-name lakes (<10 km²) which are difficult to be recognized in low-resolution RS images. Study area 2 is used to test the SR performance of our method for very small lakes on real low-resolution RS images.

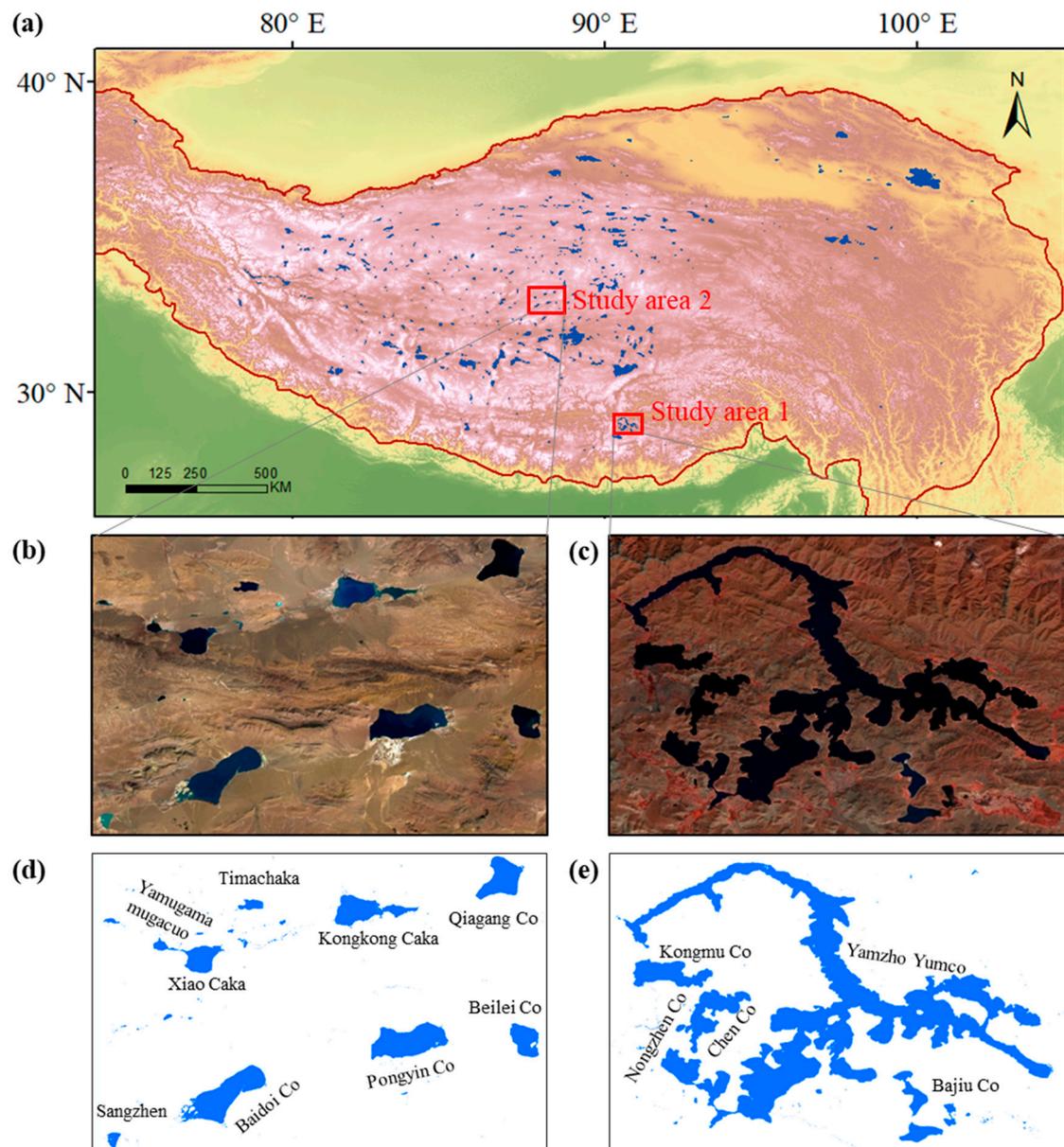


Figure 1. Study area map. (a) Tibetan Plateau (TP) area (the shapefiles of lakes are obtained from [29]), (b) Location of study area 2 shown in a color composite (R5G4B3) Landsat 8 OLI image, (c) Location of study area 1 shown in a color composite (R5G4B3) Landsat 8 OLI image, (d) Lake image derived from Landsat image of study area 2, (e) Lake image derived from Landsat image study area 1.

The Landsat-8 OLI, MODIS, and Sentinel 2 images are acquired from google earth engine (<https://earthengine.google.com>). Detailed information of the two areas is summarized in Table 1. Locations of the study areas are shown in color (R5G4B3) Landsat images in Figure 1b,c respectively. The lake images shown in Figure 1d,e are derived from the corresponding Landsat images at 30 m resolution using the NDWI method.

Table 1. Main characteristics of the two study areas and the data information.

Properties	Study area1		Study area2	
Location	28.727°–29.203°N, 90.365°–91.085°E	32.672°–33.278°N, 87.572°–88.480°E	32.675°–33.274°N, 87.574°–88.478°E	32.849°–33.277°N, 88.047°–88.497°E
Image Data	Landsat 8 OLI image	Landsat 8 OLI image	MODIS image	Sentinel 2 image
Date	October 15, 2014	October 13, 2014 August 18, 2017	October 13, 2014 August 18, 2017	August 11, 2017
Image size	2688 × 1760	3354 × 2220	202 × 135	5015 × 4767
Image resolution	30m	30m	500m	10m

2.2. Overview of the Proposed Method

Figure 2 shows the flowchart of the proposed method for generating finer-scale lakes from RS images, the core idea is to break the spatial resolution limitations of the original RS images by introducing a super-resolution method. There are two main components of the whole workflow: lake area extraction (LAE) and image SR.

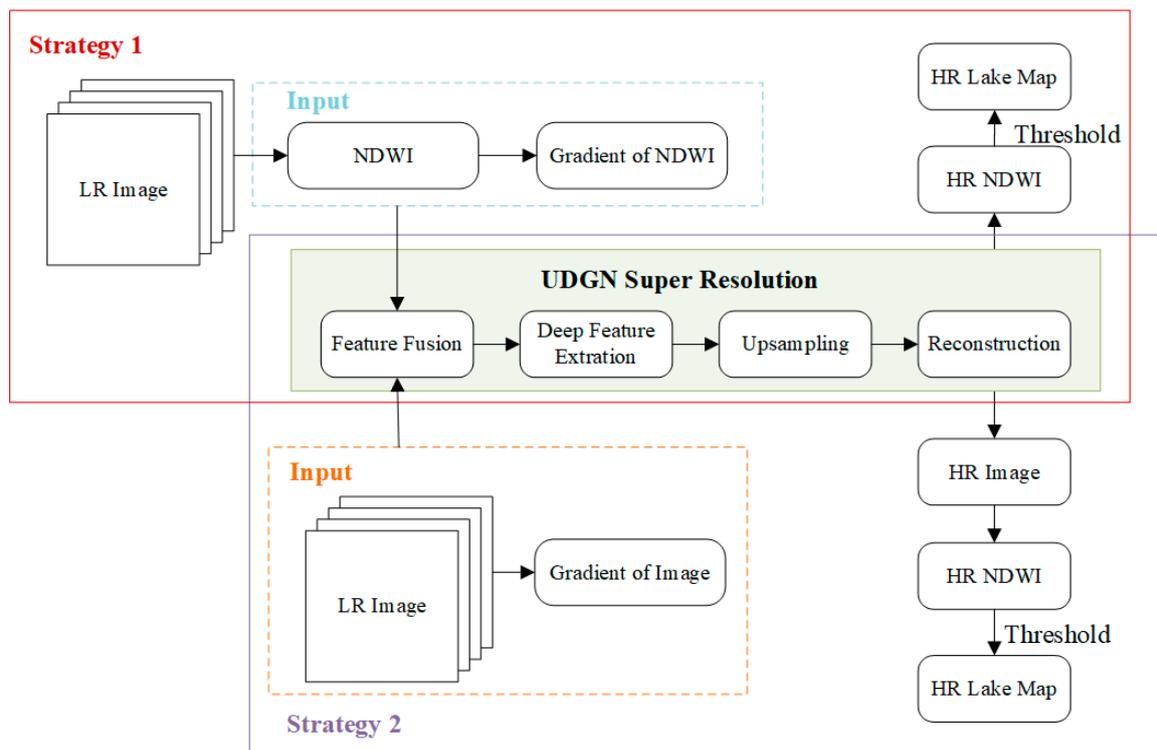


Figure 2. Flowchart of higher resolution lake area extraction (LAE) from remote sensing (RS) images. Two strategies using the unsupervised deep gradient network (UDGN) model are proposed to improve the spatial resolution of the lake area.

In LAE, the normalized difference water index (NDWI) is adopted to automatically separate water and non-water features. The NDWI has been widely used for water and lake body classification from RS images [30–32]. It is calculated according to the following equation:

$$\text{NDWI} = (\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR}) \quad (1)$$

where Green denotes the green band and NIR is the near-infrared band, and the range of the color slice with 0-1 is chosen to extract the water body boundary according to [4].

The SR method aims to improve the spatial resolutions of each input LR image. The UDG model mainly consists of four parts: feature fusion, deep feature extraction, upsampling, and reconstruction. The first part fuses the features of image gradient information and original LR images; the second part aims to extract more complex and deep features from the fused features; the third part devotes to improve the spatial resolution, and the last part finally generates the HR image.

Two strategies are proposed in this paper to improve the spatial resolution of the lake area (Figure 2). Strategy 1 is to generate the NDWI map first, then the SR model is applied to get the NDWI map with higher spatial resolution. Finally, the lake area is extracted through the threshold. Strategy 2 is to super-resolve the original multispectral RS images via the SR model, and the NDWI map is subsequently calculated from the reconstructed HR image to identify the lake areas. The biggest difference between the two strategies is the input content of the SR model. Strategy 1 takes the NDWI product as input while strategy 2 takes the original RS image. Both two strategies construct an end-to-end high spatial resolution LAE procedure to provide better support for finer lake monitoring.

In reality, we usually encounter many challenges such as lack of sufficient RS images, and it is difficult to use existing products to obtain better spatial resolution accurately. UDG is an unsupervised learning method, which can adapt to different image sizes, channel numbers, and types. With the properties of the UDG method, we do not require amounts of paired images for training, thus we can directly improve the spatial resolution of products such as the NDWI maps.

2.3. Unsupervised Super-Resolution Mechanism

Compared to supervised learning, unsupervised learning does not require paired LR-HR images for training. In this paper, inspired by [21], a new unsupervised deep learning-based SR method UDG is proposed.

In the proposed method, unsupervised learning is based on the hypothesis that the repeated occurrence of small image patches across scales of a single image is a very strong property of natural images [33,34]. It is the same for the RS images. As shown in Figure 3, the mountain and lake patches are shown to repeat many times inside the whole image. Therefore, our method relies only on the input LR image and exploits image-specific information to generate a super-resolved image.

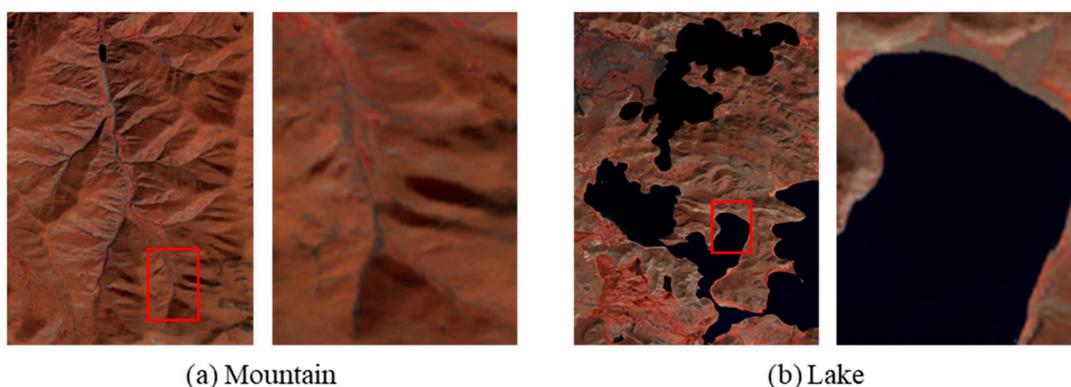


Figure 3. Examples of self-similar patterns inside a RS image.

Specifically, the unsupervised learning mechanism is shown in Figure 4. Each test image I_{test} is down-sampled first to obtain the lower-resolution image I_{LR} . Then, corresponding image patches derived from I_{LR} and I_{test} are collected as samples to train the UDG model. Last, the trained model is applied to the test image to produce the HR image I_{HR} .

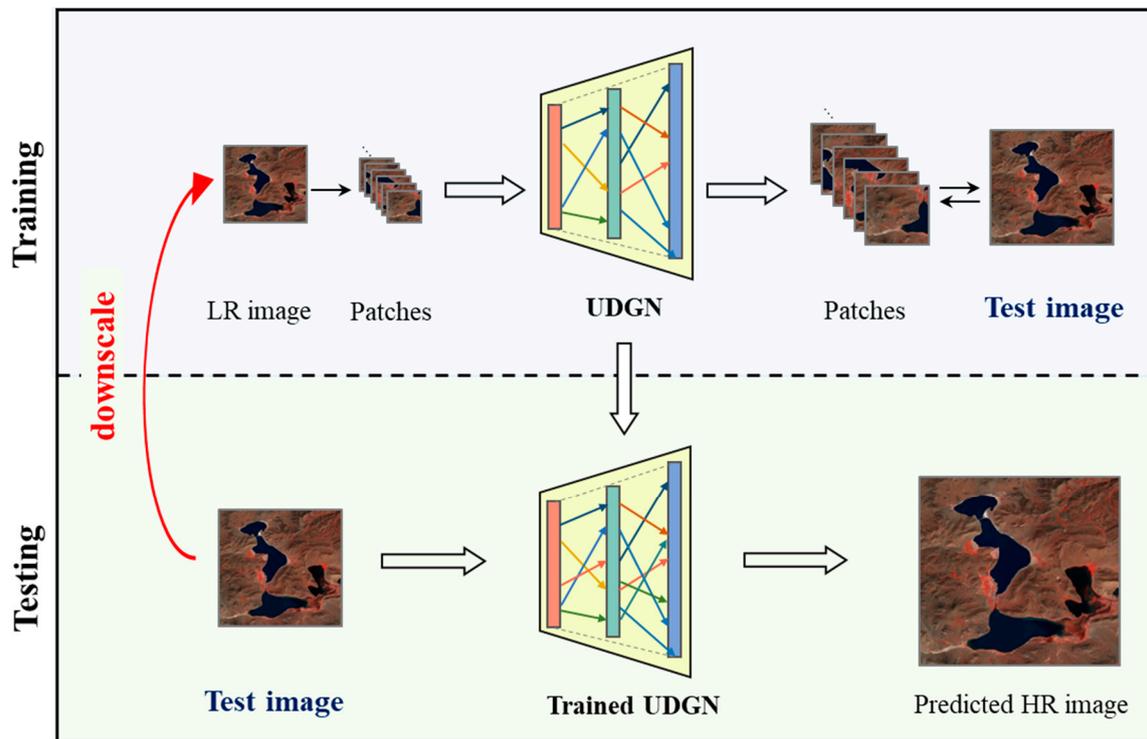


Figure 4. Unsupervised learning mechanism. An image-specific UDG model is trained on examples extracted internally, from the test image itself. The test image is firstly down-sampled to lower resolutions, and then the UDG is trained to recover the test image from its low-resolution (LR) versions. Finally, the resulting self-supervised network is applied to the test image to produce high-resolution (HR) images.

Because the learning of the model focused on a single image, it can avoid interference from other image features, image quality, noise, etc. In addition, the model can learn features more precisely and specifically. Furthermore, since our model does not require pre-training, it can adapt itself to different settings per image, such as different image sizes and different numbers of input channels. This allows us to perform SR of RS images and intermediate products (e.g., the NDWI image in this paper).

2.4. The Structure of the UDG Model

The UDG model aims to learn the cross-scale internal recurrence of image-specific information and use this information to improve the spatial resolution of each test image. Finally, extracting HR lakes from the super-resolved images. The architecture of the UDG model is shown in Figure 5. The network consists of convolution (Conv) layers, rectified linear unit (ReLU) layers, fusion layer (Fusion), element-wise-sum layers, pixel-shuffle layers, and several residual blocks (ResBlock). Conv layers are to extract low-level features and the ReLU layer is taken as an activation function for nonlinear mapping. The fusion layer is to concatenate the feature maps for feature fusion. The pixel-shuffle layer is to transform the feature maps into the size as desired for HR output.

In the feature fusion part, the gradient map of the LR image is obtained with Sobel Operator [35] first. The gradient map is important in boundary detection because images often change most quickly at the boundary between objects (Jacobs, 2005), and this information is important for lakes extraction. Then, two simple CNNs are built to preliminarily extract shallow features of both the LR image and its gradient map. The LR image can provide much low-frequency information. By fusing with high-frequency information contained in the gradient map, the integrated feature maps can retain more comprehensive details in a super-resolved HR image.

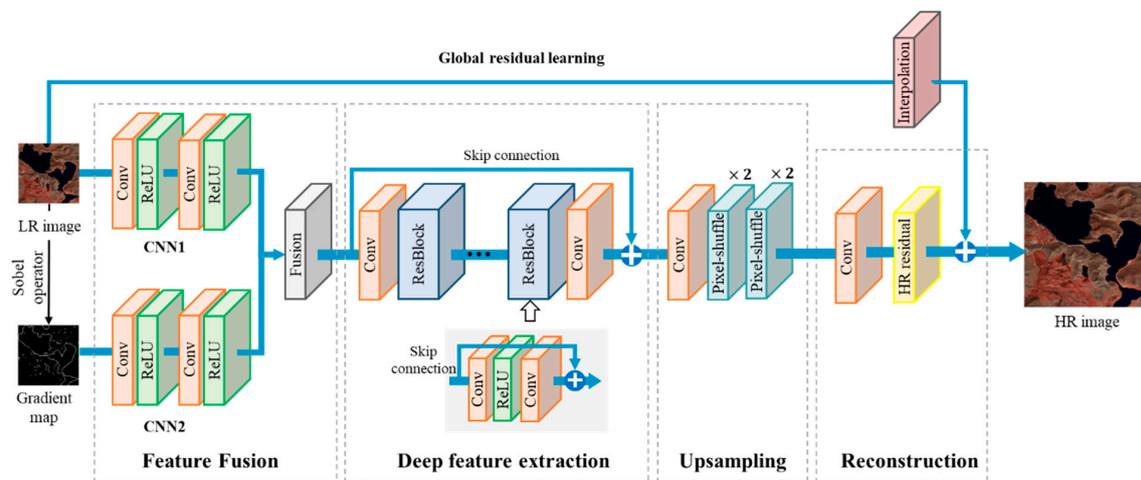


Figure 5. The architecture of the proposed UDGNet.

In the deep feature extraction part, several ResBlocks are devoted to extract high-level features and learn the complex mapping between LR and HR images. Each ResBlock consists of two Conv layers and a ReLU layer.

In the upsampling part, several pixel-shuffle layers are used to improve image size gradually. The detailed description of this kind of layer can be found in [36]. Each pixel-shuffle layer upscales two times.

In the reconstruction part, the original input LR image is interpolated to the HR size to provide global low-frequency information. By integrating the interpolated image and the HR residual, the HR image is finally obtained.

In summary, the proposed network has three main characteristics:

- I. Deep: it can efficiently extract deep features and complete multi-spectral RS-SR tasks.
- II. Geographic information preservation: By the fusion of the gradient information to enhance the original image, more geoinformation such as terrain and texture can be preserved, which provides a good foundation for subsequent LAE.
- III. Adaptive: it can super-resolve RS images/products of different image sizes and channels.

2.5. Evaluation Criteria

To evaluate the performance of the proposed method quantitatively, we adopt two groups of criteria, one group for SR performance evaluation and the other group for LAE accuracy evaluation.

Peak signal-to-noise ratio (PSNR) [37], structural similarity index (SSIM) [38], the normalized root mean square error (NRMSE) [25], and the spectral angle mapper (SAM) [39] are used to evaluate the SR performance. PSNR is measured in decibels (dB). The larger the PSNR and SSIM, the better the SR performance. The smaller the values of NRMSE and SAM, the better the SR effect.

In terms of LAE accuracy assessment, overall accuracy (OA), Kappa coefficient (KC), average producer's accuracy (APA), and average user's accuracy (AUA) are utilized. These criteria have been used in many types of research, such as water body extraction [40], urban flooding mapping [41], and wetland inundation mapping [9]. A higher value of these criteria indicates the method is of higher quality.

3. Results

3.1. Implementation Details

3.1.1. Architecture Details of UDGNet

When the upscale factor is 4, the specific settings of the components of UDGNet are listed in Table 2. The kernel size of each Conv layer is 3×3 . During the training phase, the loss function is an L1 loss, and the optimization algorithm is Adam. The learning rate is set to 0.001, and multiplied by 0.1 after 60 epochs.

Table 2. The specific architecture of UDGNet when the upscale factor is 4.

Layer	Kernel Size	Number of Kernels	Output Size	Stride	
Feature Fusion	CNN1	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	32	$n \times m$	1
	CNN2	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	32	$n \times m$	1
Deep feature extraction		$3 \times 3 \text{ Conv}$	64	$n \times m$	1
	ResBlock1	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	64	$n \times m$	1
	ResBlock2	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	64	$n \times m$	1
	ResBlock3	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	64	$n \times m$	1
	ResBlock4	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	64	$n \times m$	1
	ResBlock5	$\begin{bmatrix} 3 \times 3 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix}$	64	$n \times m$	1
		$3 \times 3 \text{ Conv}$	64	$n \times m$	1
Upsampling		$3 \times 3 \text{ Conv}$	$64 \times 4 \times 4$	$n \times m$	1
	pixel-shuffle pixel-shuffle			$2n \times 2m$ $4n \times 4m$	
Reconstruction	Interpolation			$4n \times 4m$	
	Element-wise sum			$4n \times 4m$	

3.1.2. Training Data Extraction

As illustrated in Section 2.3, for each test image, we train a specific network with training samples derived from the test image referred to [21]. Specifically, at each iteration, we take a random crop of fixed size from a randomly-selected example pair. The crop size should be smaller than the size of the input image. In this paper, the crop size is typically set to 256×256 . In addition, we use augmentation methods to generate more training examples to fully train the model. The augmentation methods include flipping, rotating, and panning the image randomly.

3.2. Results of Two Strategies

To compare the performance of the two strategies for producing a higher resolution lake area, we test our method on Landsat 8 dataset of study area 1. For strategy 1, the NDWI map is used as the input of the SR model, while for strategy 2, the original RGB RS image is used as the input. Band 5 and Band 3 are the NIR band and Green band, respectively, which are used to calculate the NDWI map as formulated in Equation (1).

The LAE results are shown in Table 3, and the visual results when the upscale factor is 4 is shown in Figure 6. In addition, Figure 7 further demonstrates the difference between the SR image and the ground-truth image. From the global aspect, Table 3 indicates that the performance of a 3 channel input is better than 1 channel. For instance, when the upscale factor is 8, the Kappa of strategy 2 is 0.9718, while that of strategy 1 is 0.9372. In addition, except for the AUA values, the OA, APA, and Kappa results of strategy 2 are all larger than strategy 1. In terms of the visual results, it is apparent that strategy 1 tends to obtain lake areas with more noises around the edges. As we can see from Figure 7b, there are many green pixels around the lakes, indicating that many land pixels are wrongly classified as lakes. In addition, using strategy 1, some small rivers (in purple circles) are super-resolved to be much bigger than the actual. In contrast, strategy 2 is able to achieve HR images with sharper and clearer lake structures and the noises are much less than strategy 1.

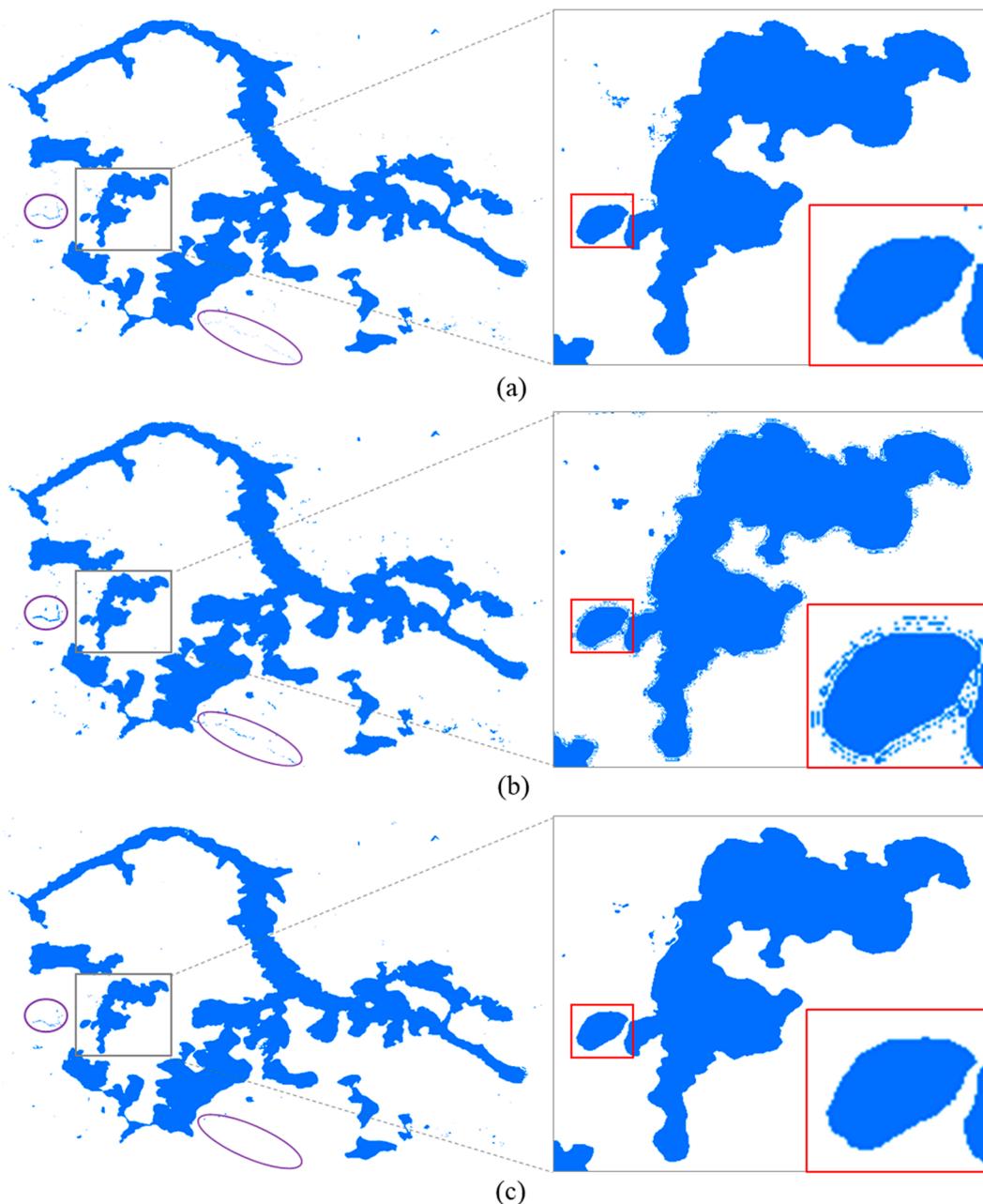
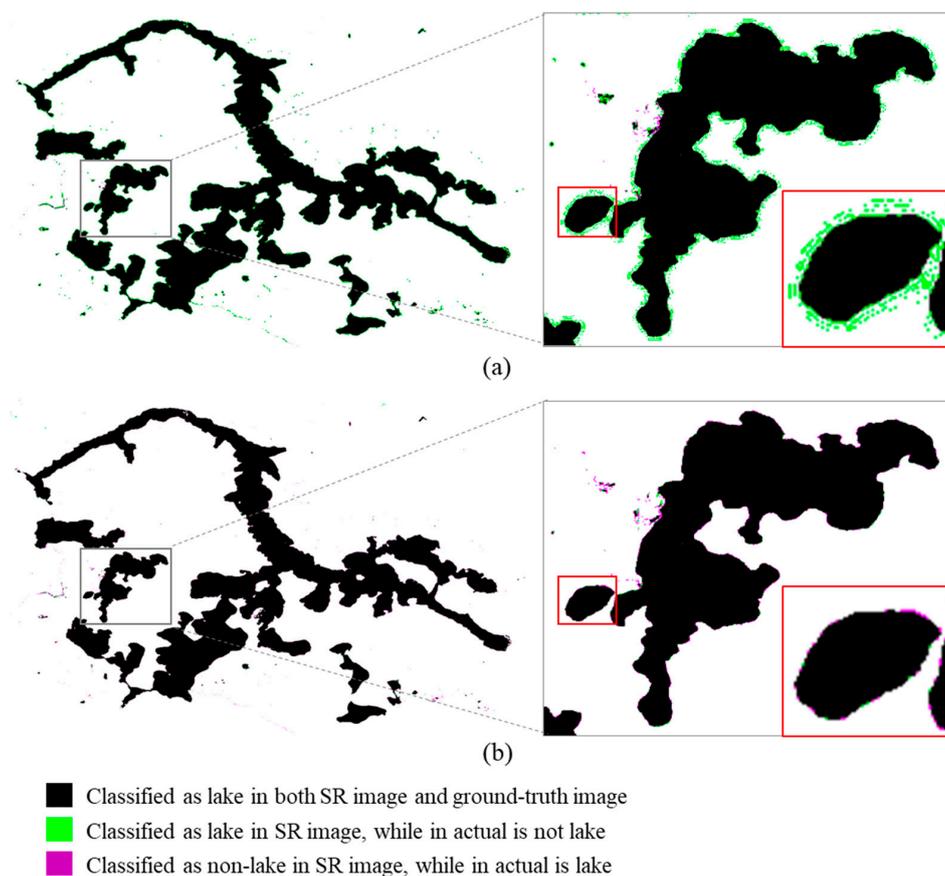


Figure 6. HR lake mapping obtained from two strategies with an upscale factor of 4. (a) ground-truth lake area, (b) lake area extracted by strategy 1, (c) lake area extracted by strategy 2.

Table 3. Detailed results of two strategies on different upscale factors.

Evaluation Criteria	Upscale Factor	Strategy 1	Strategy 2
OA	2	0.9930	0.9986
AUA	2	0.9993	0.9942
APA	2	0.9627	0.9978
kappa	2	0.9764	0.9951
OA	4	0.9902	0.9972
AUA	4	0.9987	0.9869
APA	4	0.9487	0.9972
kappa	4	0.9670	0.9904
OA	8	0.9809	0.9918
AUA	8	0.9977	0.9722
APA	8	0.9046	0.9813
kappa	8	0.9372	0.9718

**Figure 7.** Difference between SR images and ground-truth image. (a) Difference between strategy 1 and the ground-truth image, (b) Difference between strategy 2 and the ground-truth image.

The reasons for the different effects of the two strategies are as follows: first, in the NDWI calculating process, much important information is lost such as other land cover feature types, different spectral information, and topographic information around the lake. In strategy 1, the NDWI image is generated firstly and then the super-resolution is conducted directly on the NDWI image. Edge noise is easy to occur during the super-resolution since there is no more neighboring topographic and spectral information to provide as constraints. On the contrary, the calculating of NDWI is the last step of strategy 2, which can avoid the problems faced with strategy 1. In addition, using the original multi-band image as the input of the SR method, the obtained gradient map can better reflect the real

terrain condition and the relative values of the Green band and the NIR band can be well preserved in the HR image. In this way, strategy 2 is able to further improve the LAE accuracy.

In the real world, it depends on the specific demand to choose the proper strategy. For example, when the acquisition, splicing, and fusion process of original RS images is difficult. Researches can use strategy 1 to directly improve the spatial resolution of the previously generated product (e.g, lake/NDWI map). When the original image is readily available, it is recommended to use strategy 2 to obtain the lake more accurately. In addition, the fewer image channels that are input to the SR model, the less computing resources and memory they consume.

3.3. Comparison with Different SR Methods

In this section, our method is tested on the Landsat 8 dataset of study area 1. The process of all the methods is consistent with strategy 1. Firstly, the NDWI map is obtained using band 5 (NIR band) and band3 (Green band) of the Landsat 8 image. The experiments are carried out with three different upscale factors, i.e., 2, 4, and 8. Since there is no real LR-HR paired data, the original NDWI image is down-sampled using the BCI algorithm with corresponding factors to obtain LR images of different scales.

In addition, to verify the effectiveness and superiority of our method, different types of unsupervised SR methods including traditional method (i.e., BCI), machine learning methods (i.e., IBP [23] and TSR [24]) and deep learning method (i.e., ZSSR [24]) and a supervised SR method super-resolution convolutional neural network (SRCNN) [42] are used to compare the SR performance as well as the accuracy of LAE. All the methods are used considering the default settings suggested by the authors. IBP and TSR are implemented in MATLAB while others are implemented in Python. The detailed results of different upscale factors and different methods are shown in Table 4. From a global perspective, SR with larger upscale factors has worse results. Furthermore, as we can see from Table 4, our proposed method achieves the best results on all the evaluation criteria.

Table 4. Comparison results of different methods.

Upscale Factors	Evaluation Criteria	BCI	IBP	TSR	SRCNN	ZSSR	UDGN
2	OA	0.9901	0.9803	0.9871	0.9871	0.9866	0.9930
	AUA	0.9998	0.9997	0.9995	0.9983	0.9991	0.9993
	APA	0.9472	0.9005	0.9327	0.9337	0.9306	0.9627
	kappa	0.9667	0.9355	0.9570	0.9571	0.9554	0.9764
	PSNR	33.4038	34.4224	37.2856	34.9307	39.0759	39.3095
	SSIM	0.9745	0.9750	0.9819	0.9741	0.9839	0.9858
	NRMSE	0.0214	0.0190	0.0137	0.0179	0.0111	0.0108
	SAM	0.0662	0.0588	0.0423	0.0551	0.0344	0.0335
4	OA	0.9783	0.9631	0.9694	0.9801	0.9715	0.9902
	AUA	0.9996	0.9996	0.9996	0.9983	0.9988	0.9987
	APA	0.8914	0.8280	0.8534	0.9006	0.8626	0.9487
	kappa	0.9291	0.8830	0.9019	0.9348	0.9082	0.9670
	PSNR	29.4002	30.1131	33.3130	29.8395	34.5430	35.1123
	SSIM	0.9481	0.9387	0.9600	0.9481	0.9628	0.9726
	NRMSE	0.0339	0.0312	0.0216	0.0322	0.0187	0.0176
	SAM	0.1050	0.0967	0.0669	0.0986	0.0556	0.0543
8	OA	0.9562	0.9284	0.9390	0.9594	0.9386	0.9809
	AUA	0.9996	0.9995	0.9994	0.9980	0.9992	0.9977
	APA	0.8022	0.7126	0.7444	0.8150	0.7434	0.9046
	kappa	0.8631	0.7881	0.8158	0.8723	0.8148	0.9372
	PSNR	25.9517	26.5812	29.1790	23.8941	29.5749	30.0654
	SSIM	0.9081	0.8873	0.9232	0.9146	0.9234	0.9459
	NRMSE	0.0504	0.0469	0.0348	0.0639	0.0332	0.0314
	SAM	0.1565	0.1456	0.1078	0.1698	0.1029	0.0973

In terms of the SR performance, the calculated PSNR, SSIM, NRMSE, and SAM results illustrate that our method can reconstruct information from LR images better than other methods. For example, when the upscale factor is 4, the PSNR and SSIM of our method are 35.1123 dB and 0.9726, respectively, while the values of other methods are smaller than 34.6dB and 0.965, especially the BCI results, which are the worst (BCI: 29.4002dB, 0.9481).

Compared with the supervised method SRCNN, unsupervised methods are superior in image super-resolution. As we can see that when the upscale factor is 2, the PSNR and SSIM results of SRCNN are smaller than TSR. When the upscale factor is 8, the PSNR, NRMSE, and SAM results of SRCNN are even worse than BCI. This may be related to the lack of sufficient training samples. In addition, since the image size of the training samples used for supervised learning must be the same, it is necessary to cut the test image of a larger size into several small patches for super-resolution. This will affect the SR and LAE results due to a lack of global information.

Comparing different types of unsupervised methods, deep learning methods (i.e., ZSSR and UDG) are superior to BCI and machine learning methods. For example, when the upscale factor is 2, the PSNR values of deep learning methods are larger than 39dB, while PSNR values of BCI and the best machine learning method (TSR) is 33.4083dB and 37.2856dB, respectively. This is because interpolation methods do not consider the prior information of the LR images and handcrafted prior features used in machine learning methods are not sufficiently competent for the SR task.

Furthermore, the highest OA, AUA, APA, and Kappa values verify that our method has a strong ability in preserving the lake structure and edges accurately in the HR images, and further improve the spatial resolution of lakes. For example, when the upscale factor is 8, the OA values of BCI, IBP, TSR, ZSSR, and UDG are 0.9562, 0.9284, 0.9390, 0.9386, and 0.9809, respectively.

In addition to the quantitative assessments, the visual results (Figure 8) when the upscale factor is 8 are provided for a qualitative and intuitive evaluation of SR performance. Focusing on Figure 8c, the images obtained from BCI are the most blurred. This is because BCI relies heavily on the values of neighboring pixels, while other important prior information such as textures are ignored. In addition, the images super-resolved through IBP and ZSSR have obvious shadows around the edges, especially the IBP results. The existence of these edge shadows will lead to misclassification of lake margins (i.e., more land areas are classified as lake area). As for TSR results, there are regular pyramid shapes in some areas. This is related to the fact that TSR builds the internal LR-HR patch database using the scale-space pyramid of the image. These pyramid shapes, which may cross land and lakes and add more noises, will largely affect the LAE accuracy. Hence, although ZSSR and TSR can generate much sharper NDWI images than BCI, the OA values are smaller than BCI (Table 4).

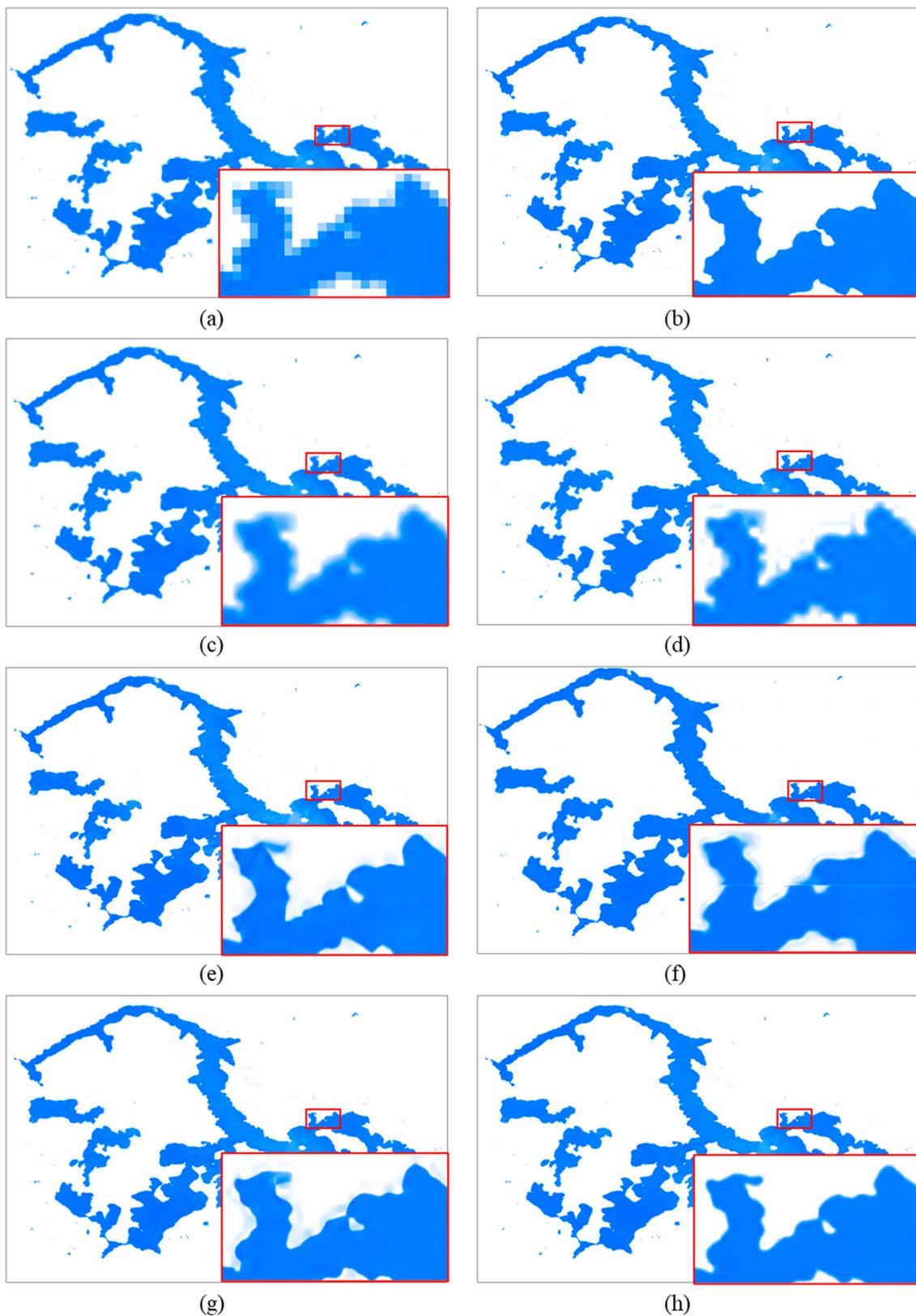


Figure 8. Normalized difference water index (NDWI) image SR results obtained by different methods with an upscale factor of 8. (a) LR NDWI image, (b) ground-truth NDWI image, (c) bicubic interpolation (BCI), (d) iterative back projection (IBP), (e) transformed self-exemplars (TSR), (f) SRCNN, (g) “zero-shot” super-resolution (ZSSR), (h) unsupervised deep gradient network (UDGN).

The proposed UDGNet is able to get high-resolution images without adding more noise. Using the deep CNN architecture with global and local residual blocks, more deep features and high-frequency information can be captured to improve the SR performance. Furthermore, by fusion of the features extracted from both the gradient map and the original test image, more geographic information details such as terrain and lake edges can remain in HR images, thus obtaining the lake area more precisely. As we can see from Figure 8h, the lake edges are sharper than other methods, and the details on the small corners are closer to reality.

As a whole, by fusing the important gradient information and learning the deep internal features of the given NDWI image, our method can significantly improve the spatial resolution of lakes, which is very important for further analysis and practical applications.

3.4. Results of Lake Extraction from MODIS Data

In this section, we verify the effectiveness of our UDGNet model in real-world scenarios. The MODIS data at study area 2 is used as the experimental dataset. Since strategy 2 outperforms strategy 1 when there are original RS images, we use the multispectral MODIS image (band 2, band 1, band 4) as the input of the UDGNet model to obtain the desired high-resolution lake area. Specifically, the MODIS image with a spatial resolution of 500m is improved to 30m using the UDGNet model. Then, the NDWI is calculated and marks the area where NDWI values are larger than 0, as the lake area.

Figure 9 shows the color images, LAE results, and provides the close-ups of some typical lakes including Pongyin Co, Timachaka, and Noname Lake. The first column presents the MODIS data, and the second column shows the predicted HR data. Landsat 8 data is displayed in the third column to reflect the actual conditions. Moreover, for the quantitative assessment of the SR performance, we roughly estimate the area of the selected three lakes by calculating the number of pixels. The results are shown in Table 5, and the estimated results are compared to the data of 2014 provided in [29].

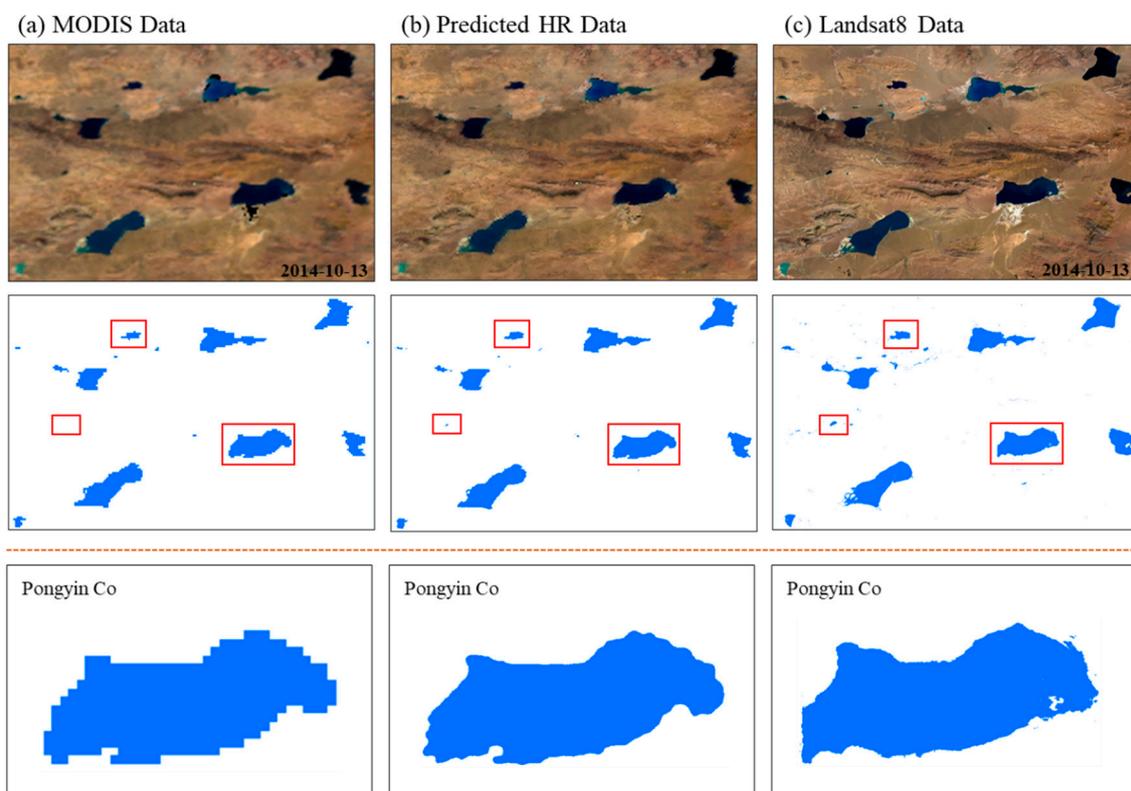


Figure 9. Cont.

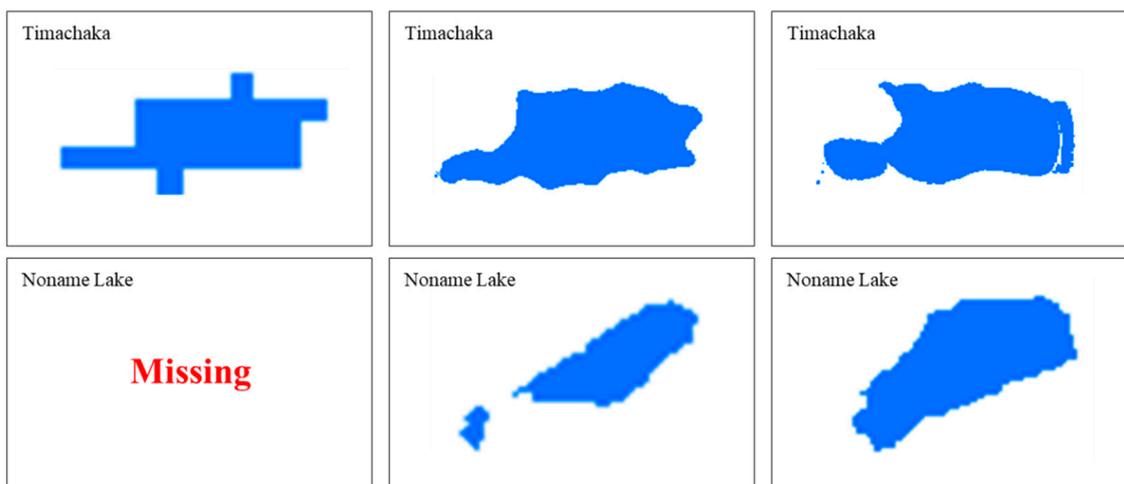


Figure 9. SR Results of real-world situations. MODIS image with an original spatial resolution of 500 m is improved to 30 m resolution. (a) MODIS data, (b) Predicted HR data, (c) Landsat 8 data.

Table 5. Area estimation results of selected lakes.

Lake Name	Pongyin Co	Timachaka	Noname Lake
MODIS image (km ²)	88.25	6.75	0
Predicted HR image (km ²)	85.3515	7.5051	0.2664
Reference data (km ²)	75.594858	7.439273	1.104846

As shown in Figure 9, the lakes derived from 500m resolution MODIS data have obvious serrated boundaries because each pixel covers a large area. It is hard to depict the actual shape of small lakes through limited pixels. By using our proposed UDG N model, more detailed structure and edge information are reconstructed, thereby, the shape of the lake is more realistic and the boundaries are smoother. In addition, the estimated area of predicted HR image is much closer to the reference data, regardless of large lakes (>10 km²) or small lakes (<10 km², <5 km²). For instance, the area of Timachaka calculated from MODIS and predicted HR data are 6.75 km² and 7.5051 km², respectively, and the reference is 7.439273 km². The results derived from the super-resolved image are much more accurate.

In addition, it is worth noting that our proposed method is able to discover small lakes from MODIS data. As illustrated in Figure 9, Noname Lake (Area: 1.104846 km²) is missing in the lake map derived from MODIS data, while the lake can be successfully extracted by improving the resolution of the image based on the UDG N model. It is of great significance in small lakes monitoring and further climate change analysis. Although there may exist differences between the predicted HR image and Landsat 8 image, it makes a big step forward to discover small lakes that cannot be figured out from the original low-resolution RS images.

4. Discussion

SR techniques can help generate a finer lake area from RS images. In this part, the practicality of the proposed method is further analyzed.

In our method, the self-learning process is done at test time. To better show the training process, the experiment is conducted taking the NDWI image of study area 1 as an example, the settings of this experiment are the same as Section 3.3. Figure 10 shows the PSNR values versus iterations when the upscale factor is 4. It can be seen that the PSNR values basically converge after about 2000 iterations. In addition, the runtime of UDG N is important in practice, and it is related to image size, upscale factors, etc. The runtime to upscale a single image of 128 × 128, 256 × 256, and 512 × 512 by 2 times is about 47, 58, and 67 seconds, respectively (on a GeForce GTX 1080 GPU).

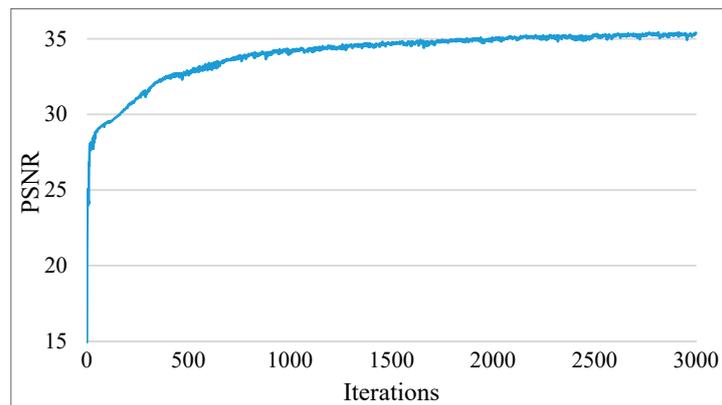


Figure 10. The Peak signal-to-noise ratio (PSNR) values versus iterations when the upscale factor is 4.

In the real world, researchers prefer to use the best available datasets such as 10 m Sentinel 2 data and mainly apply SR to the lower resolution images in order to produce a more coherent time series. Therefore, we use the UDGN model to super-resolve MODIS data (spatial resolution: 500m) and Landsat 8 data (spatial resolution: 30m) to 10m, respectively. Then, the lakes extracted from the SR images are compared with lakes extracted from a 10 m Sentinel 2 image from the same period. The results are shown in Figure 11. We can see that the lakes extracted from predicted HR data have sharper edges and details. The area of the Kongkong Caka in the original MODIS, Landsat 8, and Sentinel 2 data are 57.00 km², 60.7230 km², and 60.4132 km², respectively. Taking the Sentinel 2 data as ground-truth, the area of the Kongkong Caka extracted from the predicted HR data is closer to the ground-truth. This illustrates that the proposed method has strong practicability and can help to improve the spatial resolution of RS images and generate finer lakes.

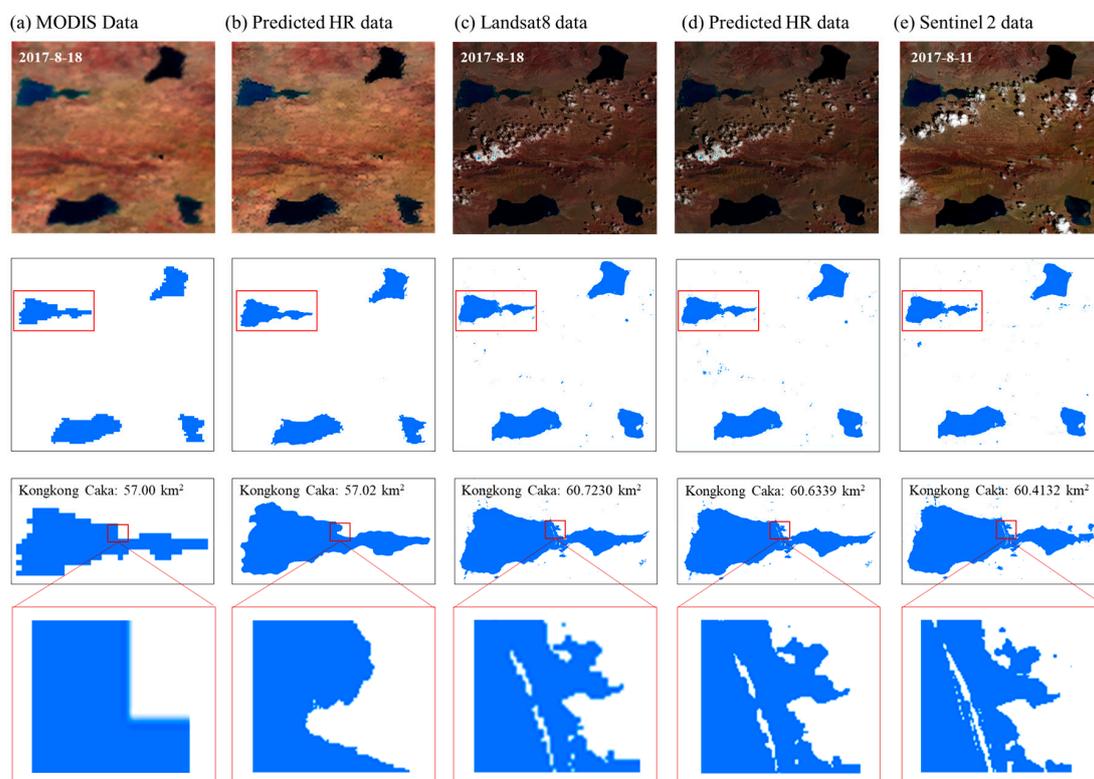


Figure 11. SR Results of MODIS image and Landsat 8 image. (a) MODIS data, (b) Predicted HR data from MODIS, (c) Landsat 8 data, (d) Predicted HR data from Landsat 8 data, (e) Sentinel 2 data.

5. Conclusions

Lake monitoring is very important for environmental and climate change studies. RS is widely used in this task, however, due to the constraints of the spatial resolution, most studies focused on large lakes, while considering small lakes is much more challenging. Therefore, in this study, we propose a new deep learning-based method UDGNet for super-resolution mapping of lakes from multispectral RS images. Unsupervised learning mechanism is exploited, which does not require a large amount of LR-HR paired samples for training. For each test image, the gradient map is obtained to retain more detailed geographical information such as edges and structures. Then, an image-specific residual network is trained at test time to improve the spatial resolution of each test image. As such, the UDGNet model can deal with different image sizes, different image channels, and different upscale factors. Based on all the above, the Landsat 8 OLI and MODIS images from two study areas from the Tibetan Plateau in China are used as experimental data. Our method is applied to improve the images with different upscale factors. The results show that our method outperforms four approaches (i.e., BCI, IBP, TSR, and ZSSR) from both visual and quantitative perspectives. It is worth noting that our method is able to discover invisible small lakes from MODIS data, which provides a way to break through the spatial resolution of RS images and better support lake studies.

Author Contributions: Conceptualization, M.Q. and Z.D.; methodology, M.Q. and L.H.; validation, M.Q., L.H., Y.G. and L.Q.; supervision, L.H., Y.G., and L.Q.; investigation, M.Q., L.H., Y.G., and L.Q.; writing—original draft preparation, M.Q.; writing—review and editing, M.Q.; project administration, Z.D., F.Z., and R.L.; resources, F.Z. and R.L.; funding acquisition, Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China under Grant 2018YFB0505000 and the National Natural Science Foundation of China (41871287).

Acknowledgments: Thanks to all the anonymous reviewers for their constructive and valuable suggestions on the earlier drafts of this manuscript.

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

References

1. O'Reilly, C.M.; Rowley, R.J.; Schneider, P.; Lenters, J.D.; McIntyre, P.B.; Kraemer, B.M. Rapid and highly variable warming of lake surface waters around the globe. *Geophys. Res. Lett.* **2015**, *42*, 10773–10781. [[CrossRef](#)]
2. 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](#)]
3. Zhang, G.; Yao, T.; Chen, W.; Zheng, G.; Shum, C.K.; Yang, K.; Piao, S.; Sheng, Y.; Yi, S.; Li, J.; et al. Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sens. Environ.* **2019**, *221*, 386–404. [[CrossRef](#)]
4. Qiao, B.; Zhu, L.; Yang, R. Temporal-spatial differences in lake water storage changes and their links to climate change throughout the Tibetan Plateau. *Remote Sens. Environ.* **2019**, *222*, 232–243. [[CrossRef](#)]
5. Lei, Y.; Yao, T.; Yang, K.; Sheng, Y.; Kleinherenbrink, M.; Yi, S.; Bird, B.W.; Zhang, X.; Zhu, L.; Zhang, G. Lake seasonality across the Tibetan Plateau and their varying relationship with regional mass changes and local hydrology. *Geophys. Res. Lett.* **2017**, *44*, 892–900. [[CrossRef](#)]
6. Song, C.; Huang, B.; Ke, L. Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. *Remote Sens. Environ.* **2013**, *135*, 25–35. [[CrossRef](#)]
7. Lei, Y.; Yang, K.; Wang, B.; Sheng, Y.; Bird, B.W.; Zhang, G.; Tian, L. Response of inland lake dynamics over the Tibetan Plateau to climate change. *Clim. Change* **2014**, *125*, 281–290. [[CrossRef](#)]
8. Zhang, G.; Yao, T.; Piao, S.; Bolch, T.; Xie, H.; Chen, D.; Gao, Y.; O'Reilly, C.M.; Shum, C.K.; Yang, K.; et al. Extensive and drastically different alpine lake changes on Asia's high plateaus during the past four decades. *Geophys. Res. Lett.* **2017**, *44*, 252–260. [[CrossRef](#)]

9. Li, L.; Chen, Y.; Xu, T.; Liu, R.; Shi, K.; Huang, C. Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm. *Remote Sens. Environ.* **2015**, *164*, 142–154. [[CrossRef](#)]
10. Xu, Z.; Chen, Z.; Yi, W.; Gui, Q.; Hou, W.; Ding, M. Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 80–90. [[CrossRef](#)]
11. Xu, Z.; Wang, X.; Chen, Z.; Xiong, D.; Ding, M.; Hou, W. Nonlocal similarity based DEM super resolution. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 48–54. [[CrossRef](#)]
12. Ma, W.; Pan, Z.; Guo, J.; Lei, B. Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive Res-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3512–3527. [[CrossRef](#)]
13. Qin, M.; Mavromatis, S.; Hu, L.; Zhang, F.; Liu, R. Remote Sensing Single-Image Resolution Improvement Using A Deep Gradient-Aware Network with Image-Specific Enhancement. *Remote Sens.* **2020**, *12*, 758. [[CrossRef](#)]
14. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 305–319. [[CrossRef](#)]
15. Xie, W.; Li, Y.; Lei, J. Hyperspectral image super-resolution using deep feature matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6055–6067. [[CrossRef](#)]
16. Han, W.; Chang, S.; Liu, D.; Yu, M.; Witbrock, M.; Huang, T.S. Image Super-Resolution via Dual-State Recurrent Networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2018**, 1654–2013. [[CrossRef](#)]
17. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projection networks for super-resolution. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2018**, 1664–1673. [[CrossRef](#)]
18. Yamanaka, J.; Kuwashima, S.; Kurita, T. Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network. *Int. Conf. Neural Inf. Process.* **2017**, 217–225. [[CrossRef](#)]
19. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 1646–1654. [[CrossRef](#)]
20. Tian, C.; Xu, Y.; Fei, L.; Yan, K. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Int.* **2018**, 1–23. [[CrossRef](#)]
21. Shocher, A.; Cohen, N.; Irani, M.; Math, A. “Zero-Shot” super-resolution using deep internal learning. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2018**. [[CrossRef](#)]
22. Sun, J.; Xu, Z.; Shum, H.-Y. Image super-resolution using gradient profile prior. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 23–28 June 2008; pp. 1–8.
23. Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP Graph. Model. image Process.* **1991**, *53*, 231–239. [[CrossRef](#)]
24. Huang, J.-B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2015**, 5197–5206. [[CrossRef](#)]
25. Haut, J.M.; Fernandez-Beltran, R.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Pla, F. A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2018**, *11*, 6792–6810. [[CrossRef](#)]
26. Bulat, A.; Yang, J.; Tzimiropoulos, G. To learn image super-resolution, use a GAN to learn how to do image degradation first. In Proceedings of the The European Conference on Computer Vision (ECCV), Munich, Germany, 13–18 September 2018.
27. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Work.* **2018**, 814–823. [[CrossRef](#)]
28. Yasuhiro, S. Journal of Geophysical Research: Preface. *J. Geophys. Res. Atmos.* **2015**, *120*, 4764–4782.
29. Wan, W.; Long, D.; Hong, Y.; Ma, Y.; Yuan, Y.; Xiao, P.; Duan, H.; Han, Z.; Gu, X. A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014. *Sci. Data* **2016**, *3*, 1–13. [[CrossRef](#)] [[PubMed](#)]
30. Tao, S.; Fang, J.; Zhao, X.; Zhao, S.; Shen, H.; Hu, H.; Tang, Z.; Wang, Z. Rapid loss of lakes on the Mongolian Plateau. **2015**, *112*, 2281–2286. [[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**, *14*, 3025–3033. [[CrossRef](#)]
32. Mcfeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]

33. Zontak, M.; Irani, M. Internal Statistics of a Single Natural Image. *CVPR* **2011**, 977–984. [[CrossRef](#)]
34. Glasner, D.; Bagon, S.; Irani, M. Super-Resolution from a Single Image. In Proceedings of the 2009 IEEE 12th international conference on computer vision, Kyoto, Japan, 29 September–2 October 2009; pp. 349–356.
35. Sobel, I. An Isotropic 3x3 Image Gradient Operator. *Present. Stanford A.I. Proj.* **1968**, 271–272.
36. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 1874–1883. [[CrossRef](#)]
37. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [[CrossRef](#)]
38. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
39. Yuhas, R.H.; Goetz, A.F.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. *Proc. Summ. Annu. JPL Airborne Geosci. Work.* **1992**, 147–149.
40. Wang, X.; Ling, F.; Yao, H.; Liu, Y.; Xu, S. Unsupervised Sub-Pixel Water Body Mapping with Sentinel-3 OLCI Image. *Remote Sens.* **2019**, *11*, 327. [[CrossRef](#)]
41. Li, L.; Xu, T.; Chen, Y. Improved urban flooding mapping from remote sensing images using generalized regression neural network-based super-resolution algorithm. *Remote Sens.* **2016**, *8*, 625. [[CrossRef](#)]
42. Dong, C.; Loy, C.C.; He, K.; Tang, X. *Learning a Deep Convolutional Network for Image Super-Resolution*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).