

Article

Road Extraction in Mountainous Regions from High-Resolution Images Based on DSDNet and Terrain Optimization

Zeyu Xu ^{1,2} , Zhanfeng Shen ^{1,3,*}, Yang Li ^{2,4} , Liegang Xia ⁵, Haoyu Wang ^{1,2}, Shuo Li ^{1,3}, Shuhui Jiao ^{1,2} and Yating Lei ^{1,2}

- ¹ National Engineering Research Center for Geomatics, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; xuzeyu@aircas.ac.cn (Z.X.); wanghaoyu@aircas.ac.cn (H.W.); lishuo191@mails.ucas.ac.cn (S.L.); jiaosh@aircas.ac.cn (S.J.); yatinglei@mail.ynu.edu.cn (Y.L.)
 - ² School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China; liyang194@mails.ucas.ac.cn
 - ³ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
 - ⁴ State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
 - ⁵ College Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310014, China; xialg@zjut.edu.cn
- * Correspondence: shenzf@aircas.ac.cn

Abstract: High-quality road network information plays a vital role in regional economic development, disaster emergency management and land planning. To date, studies have primarily focused on sampling flat urban roads, while fewer have paid attention to road extraction in mountainous regions. Compared with road extraction in flat regions, road extraction in mountainous regions suffers more interference, due to shadows caused by mountains and road-like terrain. Furthermore, there are more practical problems involved when researching an entire region rather than at the sample level. To address the difficulties outlined regarding mountain road extraction, this paper takes Jiuzhaigou county in China as an example and studies road extraction in practical applications. Based on deep learning methods, we used a multistage optimization method to improve the extraction effect. First, we used the contrast limited adaptive histogram equalization (CLAHE) algorithm to attenuate the influence of mountain shadows and improve the quality of the image. Then the road was extracted by the improved DSDNet network. Finally, the terrain constraint method is used to reduce the false detection problem caused by the terrain factor, and after that the final road extraction result is obtained. To evaluate the effect of road extraction comprehensively, we used multiple data sources (i.e., points, raster and OpenStreetMap data) in different evaluation schemes to verify the accuracy of the road extraction results. The accuracy of our method for the three schemes was 0.8631, 0.8558 and 0.8801, which is higher than other methods have obtained. The results show that our method can effectively solve the interference of shadow and terrain encountered in road extraction over mountainous regions, significantly improving the effect of road extraction.

Keywords: road extraction; deep learning; DSDNet; CLAHE; terrain constraints



Citation: Xu, Z.; Shen, Z.; Li, Y.; Xia, L.; Wang, H.; Li, S.; Jiao, S.; Lei, Y. Road Extraction in Mountainous Regions from High-Resolution Images Based on DSDNet and Terrain Optimization. *Remote Sens.* **2021**, *13*, 90. <https://doi.org/10.3390/rs13010090>

Received: 17 November 2020

Accepted: 24 December 2020

Published: 29 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

High-quality road network information plays an important role in many practical applications [1]. In urban regions the extraction of road network information can be used to make maps and for urban planning. In mountainous regions, the extraction of road network information can be used to make travel route planning adjustments and informs emergency decision-making for disasters such as earthquakes and floods. Road extraction from remote sensing imagery gained considerable attention, while this task remains challenging owing to irregular and complex road sections and structures [2].

With the recent development in artificial intelligence, especially deep learning technology, new ideas have been provided for road extraction [3]. Deep learning was developed

from research of neural networks and has made significant advancements in computer vision [4,5] and other fields. It has shown great advantages in image scene classification [6,7], target detection [8,9] and semantic segmentation [10–12]; the latter is a useful tool for road extraction [3] so we use it here as the basis for our method. However, from the perspective of practical applications, there are many special issues that need to be studied. The problems associated with extracting roads from complete mountainous regions are summarized as follows:

(1) The shadowed area of mountains will cause transmission change of the spectrum. Unlike urban architectural shadows, there are large areas of shadow in mountainous regions.

(2) Mountainous regions are cloudy on most days of the year so it is often difficult to guarantee a high-quality data source for mountainous regions. There are higher requirements for feature extraction and recognition.

(3) Due to mountain topography, the spectrum and shape of valleys and the bare regions where water once flowed are very similar to the roads, which can cause false detection.

The above problems are based on the complete regional analysis of Jiuzhaigou county. If we only selected some samples, some of these problems may be ignored. Therefore, it is necessary to study the road extraction of a complete region. Here, we address the problems discussed as follows:

(1) We enhance the image quality through the contrast limited adaptive histogram equalization (CLAHE) algorithm to alleviate the problems of mountain shadows and low-quality images.

(2) We propose a new network to improve the extraction ability of the neural network by using the squeeze-and-excitation module and integrating dense upsampling convolution into the encoder–decoder model. We also improve the learning rate decay method and activation function.

(3) The interference caused by the complex terrain is eliminated using a post-processing method. We use the least squares method to fit the road direction; then we calculate the road-gradient angle and the road-direction slope through topographic analysis to constrain the road extraction results.

This article uses a variety of precision evaluation schemes to study mountain road extraction from the perspective of practical applications, including cross-validation using raster data, validation using point data, and validation using OSM (OpenStreetMap, <https://www.openstreetmap.org/>) data.

The remainder of this article is organized as follows: Section 2 reviews the relevant literature on road extraction; Section 3 introduces the principle of the method proposed in this article; Section 4 introduces the research region and experimental plan of this article; Section 5 introduces the basic accuracy evaluation indicators and three accuracy evaluation schemes used in this article; Section 6 introduces the experimental results and related discussions and Section 7 summarizes our results and discusses future research directions.

2. Related Work

The mountain road network can be drawn by manual editing and field inspection. However, these methods consume a lot of manpower and financial resources, and it is difficult to update on a large scale over a short time. Many automated methods have been used to extract road information such as spectral feature-based method, object-oriented method, shallow machine learning method, deep learning method and so on. Previous studies have classified existing road extraction algorithms from different point of views [13–15]. Considering the particularity of deep learning methods, in this article, we divided the existing methods into non-deep learning methods and deep learning methods.

2.1. Non-Deep Learning Methods

There are some useful algorithms used in road extraction that do not come from learning methods. Gruen et al. [16] proposed a semiautomatic road extraction scheme that used wavelet decomposition and a model-driven linear feature extraction algorithm. It is robust in gaps and long roads but the extraction accuracy needs to be improved. Hinz et al. [17] proposed a model that comprised explicit knowledge about geometry, radiometry, topology and context. They also considered the influence of some mountain terrain factors. Their proposed approach is suited for images with a resolution from 0.2 to 0.5 m, but there are still obvious discontinuities in their results and their model is too complex, which was only tested on high-quality images in small areas. Zhang et al. [14] proposed a method combining stroke width transformation and mean drift. This method performs image segmentation through mean shift and extracts roads through stroke width transformation. When the road features are obvious, this method can achieve good results. However, the applicability of the algorithm is greatly affected by the images of surrounding objects. Zhao et al. [18] proposed a method based on the marked point process with local structure constraints. This method uses a random point process to define the position, and uses line segments as the mark to define the geometric structure. By using the characteristics of the road, the road network is extracted by combining Bayesian theory and the reversible jump Markov Chain Monte Carlo algorithm. This method has good extraction accuracy, but it is difficult to extract the contour information of the road. Peng et al. [19] used an adaptive spatial filtering algorithm to control the watershed algorithm for segmentation. They extracted roads through features such as geometry and texture. This method can extract roads with high accuracy, but the effect of contour edge information extraction is problematic.

The shallow machine learning methods were also used in road extraction. Soni et al. [15] proposed a supervised multilevel framework based on least squares support vector machine (LS-SVM), mathematical morphology and road shape features to extract road networks from remote sensing images. This method uses the LS-SVM method to segment the image into road and non-road regions, and then uses the morphological and shape features to extract non-road objects. This method has good results with high-quality images, but does not perform as well in regions with complex spectra and complex road junctions.

In the above methods, there are relatively good results for road extraction, but issues including the demand of high image quality, poor contour boundary effect and complicated feature design and extraction are obvious.

2.2. Deep Learning Methods

The rise of deep learning technology has had a strong impact in the field of image processing. Many scholars have conducted in-depth research on the application of deep learning in road extraction [20,21] and different effective neural network structures have been applied and improved for road extraction.

The use of CNN has shown good results in road extraction. Bastani et al. [22] proposed the RoadTracer method, which uses an iterative search process guided by CNN-based decision-making functions to directly derive a road network map from the output of CNN with a relatively small error rate. Qi et al. [23] combined the structure of multiscale convolution and attention mechanism with the LinkNet network and obtained the ATD-LinkNet, which can effectively use spatial and semantic information in remote sensing images. In view of the low quality of road boundary extraction and the discontinuity of road extraction, Zhou et al. [1] proposed a boundary and topological-aware road extraction network (BT-RoadNet), a coarse-to-fine architecture composed of a coarse map prediction module (CMPM) and a fine map prediction module (FMPM); this method handles interruptions caused by shadows and occlusions well.

To solve the problem of limited quantity of data, some scholars applied generative adversarial networks (GANs) [24] to road extraction. Zhang et al. [15] proposed a method

based on the generative adversarial network, which showed better performance than other methods on the Massachusetts roads dataset. GAN can also be used to estimate the roads covered by trees or shadows. Zhang et al. [25] proposed a multi-supervised generative adversarial network (MsGAN), which learns the spectral and topology features. Costea et al. [26] proposed a new dual-hop generative adversarial network (DH-GAN) and applied a smoothing-based graph optimization to road extraction. This method performed well on a large dataset with European roads.

Deep transfer learning [27] is also used for road extraction. Transfer learning refers to a learning process that uses the similarity between data, tasks or models to apply the information learned in the previous domain to a new domain [28]. Deep transfer learning is the method that uses deep models such as deep neural networks for transfer learning [29]. Senthilnath et al. [30] used deep transfer learning and ensemble classifier to extract roads from UAV (unmanned aerial vehicle) imagery. He et al. [31] trained a U-Net-based baseline network on a large-scale remote sensing image dataset and used the cross-mode training data to fine-tune the first two convolutional layers of the pretrained network to achieve adaptation to the local features of the cross-mode data. These methods do not usually require large amounts of data to achieve good extraction results, so they may be applicable to mountain road extraction.

In addition, studies have used different data sources. Zhang et al. [32] adjusted the U-net network and used sentinel-1 SAR (Synthetic Aperture Radar) imagery's dual-polarimetric (VV and VH) to conduct road extraction research and improve the accuracy of road extraction. However, extraction using high-resolution data may be somewhat different in feature extraction. Henry et al. [33] combined CNN and a tolerance rule for a spatially small mistake to reach an effective solution and perform road extraction from SAR images. However, high resolution SAR data are not easily obtained and are prone to high noise interference.

Most of the existing research only studied road extraction from sample locations. However, as described in Section 1 of this article, there will be some specific problems for the extraction of a road network at the region level. It is important to pay attention to the practical application of methods and solve problems for entire regions. Some scholars have studied land cover mapping based on deep learning at a large scale [34,35], but it lacked specificity for road extraction. Salberg et al. [36] used a fully convolutional network for large-scale mapping of small roads, but they used lidar images, which are not easy to obtain. In addition, road extraction in mountainous regions is a context that has rarely been studied. Although some studies have considered terrain factors [25], they only focused on common factors like occlusion and shadows but did not consider some interference caused by the spatial and spectral characteristics of mountains. In the study by Courtial et al. [37] on mountain roads generalization, some interference can be resolved. However, some details will be ignored, which is not applicable in some situations such as earthquake emergency. Therefore, it is useful to study the road extraction of the complete mountainous region. This paper takes Jiuzhaigou county as an example and improves road extraction effect from the perspective of practical application.

3. Materials and Methods

3.1. Overall Process of Road Extraction

The VHR (very high-resolution image) data used in this study are Google Earth images. The images are displayed at different spatial separations from the Earth level and we used the images with a spatial resolution of 0.6 m. These images are in the RGB color model with 8-bit per color. Google Earth images are widely used [38], but unlike some images with more color bands and larger dynamic range, these usually require better feature extraction and image processing methods.

Starting from practical applications, we addressed the issue of road extraction from a complete mountainous region from different aspects. The method described in this paper can be divided into the following three aspects:

- Preprocessing: We used the CLAHE algorithm to perform targeted preprocessing for mountain road extraction.
- Network: We proposed DSDNet with optimizations of the existing network model.
- Postprocessing: We calculated some indicators to constrain the extraction results according to the characteristics of the mountain terrain.

Figure 1 shows the overall flow chart of road extraction. In some studies, some interrupted roads are connected by topology [1] or other similar methods [22]. However, these kinds of methods are not very suitable for this study because road interruptions are often a reality, especially after earthquakes, floods or other natural disasters. These interrupted regions are also important information. Therefore, the method in this article will not use additional methods to connect all the suspected interrupted regions. It is worth noting that the CLAHE method and DSDNet network can optimize the road interruption problems of vehicles, shadows and thin clouds through their effective feature extraction capabilities.

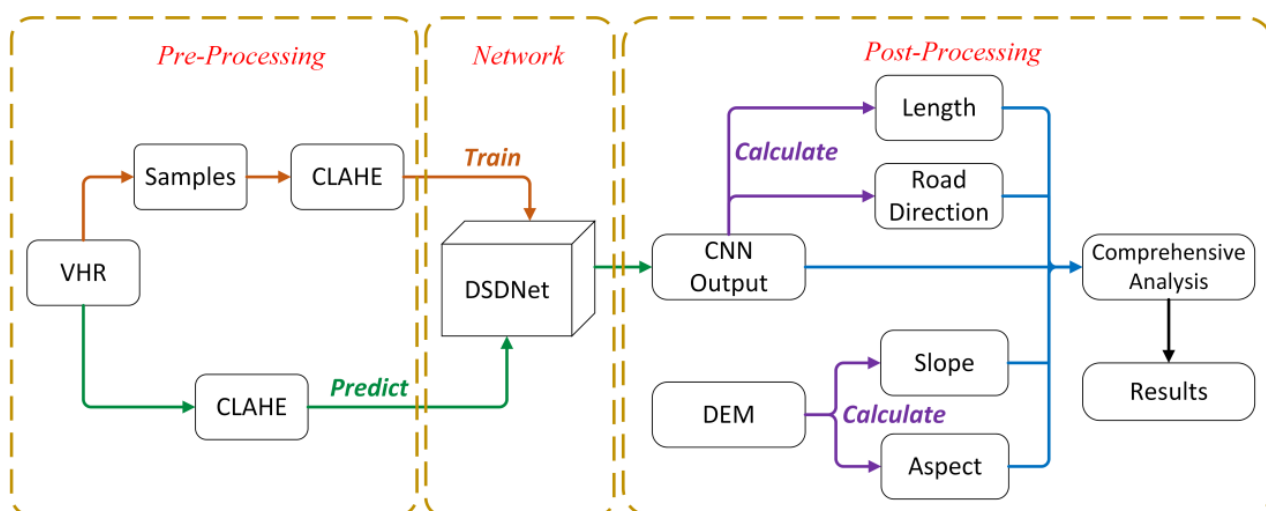


Figure 1. Overall flow chart of road extraction.

3.2. CLAHE Algorithm

An image histogram reflects the statistics of the different gray levels of the image. Through the histogram equalization (HE), the brightness can be better distributed on the histogram, so the contrast of the image can be enhanced. However, when there is a place in the image that is obviously brighter or darker than other regions, ordinary histogram equalization algorithms cannot describe the details of the place. The adaptive histogram equalization (AHE) algorithm achieves the effect of expanding the local contrast and displaying the details of the smooth region by performing histogram equalization in a rectangular region around the pixel being processed. However, there is still some noise in images obtained by the AHE algorithm. The CLAHE algorithm can deal with this problem by limiting the contrast improvement of the AHE algorithm. The contrast enlargement around the specified pixel value is mainly determined by the slope of the transformation function. This slope is proportional to the slope of the cumulative histogram of the field. CLAHE cuts the histogram with a predefined threshold before calculating the CDF (cumulative distribution function) to limit amplitude.

The main steps of the CLAHE algorithm are as follows:

- (1) Extend the image boundary so that it can be just divided into several sub-blocks.
- (2) Divide the image into blocks, take the block as the unit, first calculate the histogram, then trim the histogram and carry out equalization. For each gray level of each sub-block's histogram, use the preset limit value to limit and count the number of pixels that exceed the limit in the entire histogram.

- (3) Go through each image block and measure the linear difference between blocks.
- (4) Carry out layer filter blending operation with the original image.

We used HE and CLAHE algorithms for enhancement processing on high-resolution remote sensing images; example results are shown in Figure 2. Both the CLAHE and HE algorithms can enhance the display effect of the original image to a certain extent. However, while the HE algorithm enhances the display effect, it also interferes with road extraction. The spectral information for the road and the ground in the image obtained by the HE algorithm in the upper right corner was very similar, which is prone to misdetection during feature extraction. In the image enhanced by the CLAHE algorithm, the road in the upper right corner was still easy to distinguish.

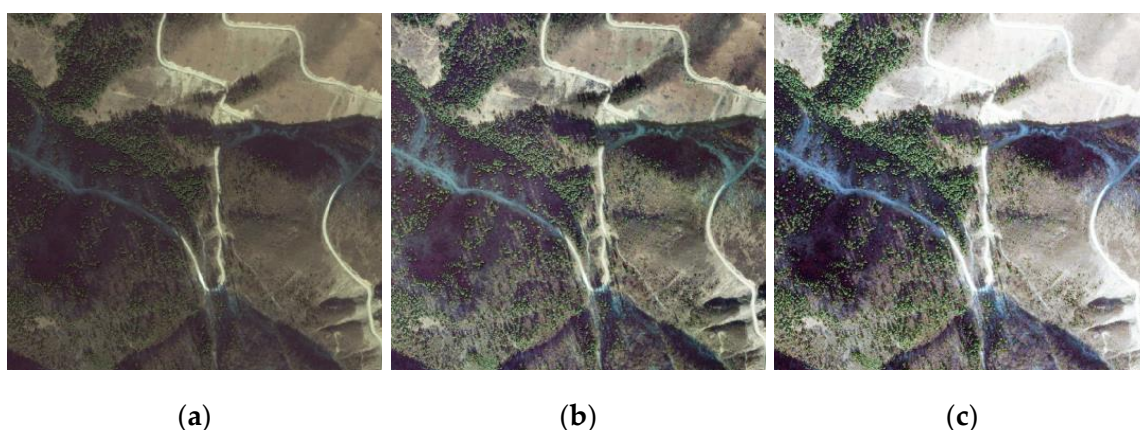


Figure 2. Diagram of image preprocessing: (a) Original image; (b) image processed by contrast limited adaptive histogram equalization (CLAHE) and (c) image processed by histogram equalization (HE).

3.3. Network Structure

Based on a variety of excellent neural network structures, we proposed the DSDNet network structure (network with dilated convolution, SE module and dense upsampling convolution). The DSDNet network model structure is shown in Figure 3. The main features of this structure are as follows:

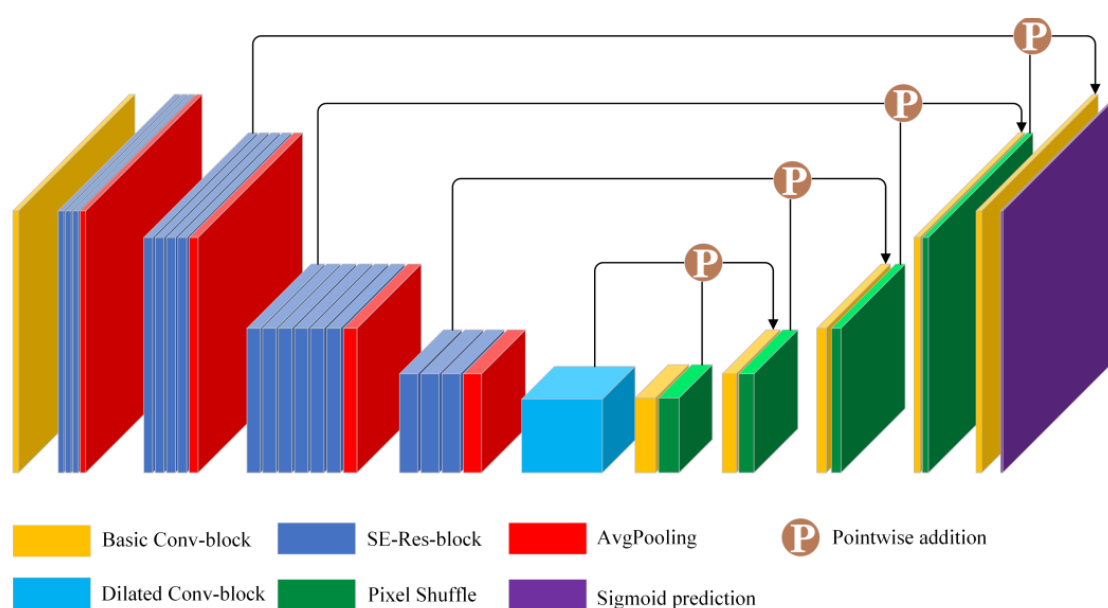


Figure 3. Structure diagram of DSDNet.

- (1) DSDNet uses encoder–decoder structure as its basis.
- (2) The encoder is improved from the D-LinkNet network [39]. DSDNet uses the ResNet [9] pretraining model and the empty convolution pooling module. Due to the channel characteristics, our network added the SE module to the base of each ResNet network to better analyze the number of channels.
- (3) The decoder adopts a dense upsampling method and integrates it into the encoder–decoder model. It is realized by a convolutional layer and a pixelshuffle layer.
- (4) DSDNet uses Leaky ReLU as the activation function.
- (5) DSDNet uses a new method with two control variables to optimize the learning rate.

DSDNet uses the encoder–decoder structure, which is used in many other effective networks, such as SegNet [12], U-Net [13], LinkNet [14], DeepLab v3+ [40], etc. For the encoder, multiple levels of features of the image were obtained through multilayer convolution, and for the decoder the extraction results were restored to the original image size in a certain way. The encoder and decoder in DSDNet were connected through pointwise addition. Through this method, the position and contour information of the extracted object can be preserved while extracting the categorical information from the image.

In the encoder, DSDNet retains the part of the ResNet network, so through this kind of transfer learning method we can quickly train a better model on our dataset. In order to extract information effectively, DSDNet adds an SE module in the encoder section. For convolution operations, most work is focused on improving the receptive field, that is, to fuse more feature fusion spatially, or to extract multiscale spatial information such as the multibranch structure of the inception network [8]. The convolution operation basically defaults to fusing all channels of the input feature map for feature fusion of channel dimensions. The squeeze-and-excitation (SE) module [41] focuses on the relationship between channels, hoping that the model can automatically learn the importance of different channel features. The SE module performs the squeeze operation on the feature map to obtain the channel-level global features, and then performs the excitation operation on the global features to learn the relationship between each channel and obtains the weights of different channels. It then multiplies the channel features by the original features to obtain the final features. Since there are many mountainous regions with poor-quality imagery, the SE module can help to extract features well. We adjusted the method of reading the pretrained model so that the network could partly read the pretrained parameters of ResNet while adding an SE module.

In the decoder, DSDNet uses a modified dense connection upsampling method. Wang et al. [42] designed dense upsampling convolution (DUC) for upsampling. This method compensates for the loss of length and width through channel dimensions. DSDNet uses a convolutional layer and a pixelshuffle layer to achieve similar functions. Other differences are that DSDNet uses multilayer upsampling and connects each layer to the encoder. The pixelshuffle algorithm was first used for super-resolution reconstruction [43]. Unlike the bilinear interpolation method, this algorithm uses subpixel convolution to achieve feature map magnification and upsampling (Figure 4). The process can be trained to achieve better results.

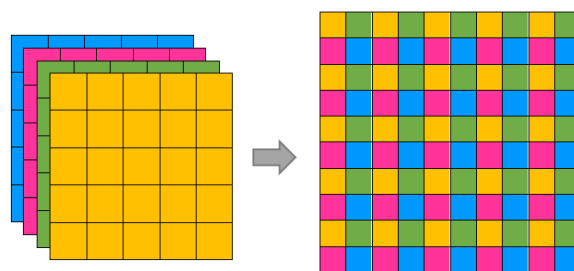


Figure 4. Schematic diagram of subpixel convolution.

DSDNet uses two control variables to optimize the learning rate (Figure 5). Both variables are related to the loss of each epoch. In one epoch, if the loss exceeds the previous minimum loss, the values of the two variables will be increased by 1. If not exceeded, variable 1 is reset to 0 and variable 2 remains unchanged. We set dynamic thresholds for the two variables, and the thresholds are determined according to the magnitude of the loss. If one of the variables reaches the threshold, both variables will be reset to 0 and the learning rate will be updated (multiplied by the change factor). In addition, we set a minimum learning rate, and when the learning rate reaches 0.000002, it will not be updated. In addition, we used Leaky ReLU to replace the original ReLU (rectified linear unit) as the activation function.

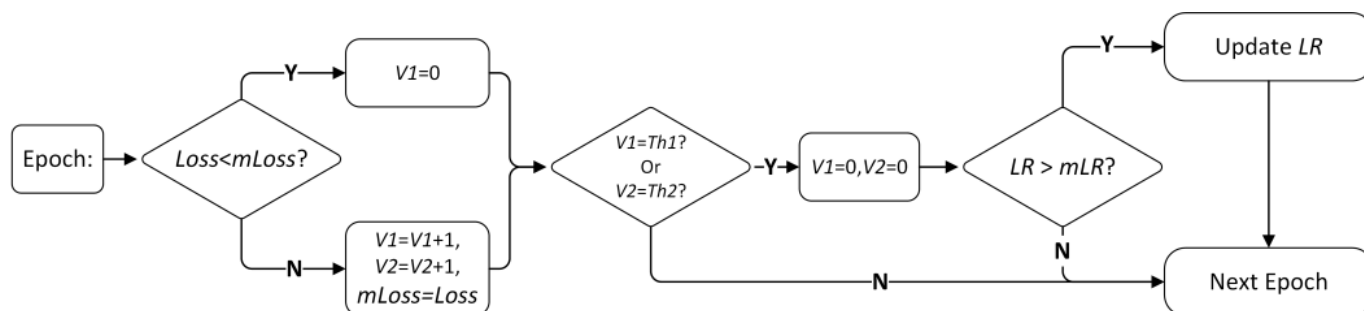


Figure 5. Method of optimizing the learning rate: $V1$ and $V2$ represent the two variables, $Th1$ and $Th2$ represent the thresholds, LR represents the learning rate and $mLoss$ and mLR represent the minimum loss and minimum learning rate, respectively.

3.4. Terrain Constraints Processing

Due to the terrain of mountainous regions there are often streams in the valley. After these creeks dry up, bare riverbeds often have a similar appearance to roads. Even if there are no riverbeds, the valley itself is very similar to a road. After the processing of CLAHE and DSDNet, the false detection caused by the special terrain can be partly solved. We went on to analyze the remaining false detection regions:

- The length of bare riverbeds and valley lines is usually very short.
- The directions of these bare riverbeds are sometimes similar to the slope gradient, but the roads are not. We proposed the road-gradient angle to represent the angle of road direction and the steepest direction.
- The slope in the road direction is usually small but the bare riverbeds and valley lines are not. We used road-direction slope to represent the slope at the road direction.

Each individual condition outlined above is not enough to distinguish between roads and non-roads. For example, some roads are relatively short and some special roads in mountainous regions may also follow directions of slope gradient. However, if we consider the sum of these conditions, we can distinguish roads better. Therefore, we could consider these three conditions at the same time by setting appropriate thresholds, and only the predictions that meet the conditions with shorter length, smaller road-gradient angles and smaller road-direction slopes will be judged as non-roads. It should be noted that we did not consider mountain roads, which are too narrow. These mountain roads can be almost perpendicular to the ground, but they are different from the road typical roads. In addition, in mountainous regions, there are many places to walk, and we cannot use all these places as road extraction objects. Therefore, we set three thresholds: The maximum length of the road, the maximum value of the road-gradient angle and the maximum value of road-direction slope. The predictions within these three thresholds are regarded as non-roads and will be removed.

Figure 6 shows a schematic diagram of mountain terrain variables. The length of the road is easy to calculate and then we fit the road to a straight line by the least square method and calculated the road direction according to the direction of the straight line. Combined with aspect, which was calculated from the DEM (Digital Elevation Model), the

road-slope angle was calculated. Then, we calculated the road-direction slope by using the straight line of road and the DEM data.

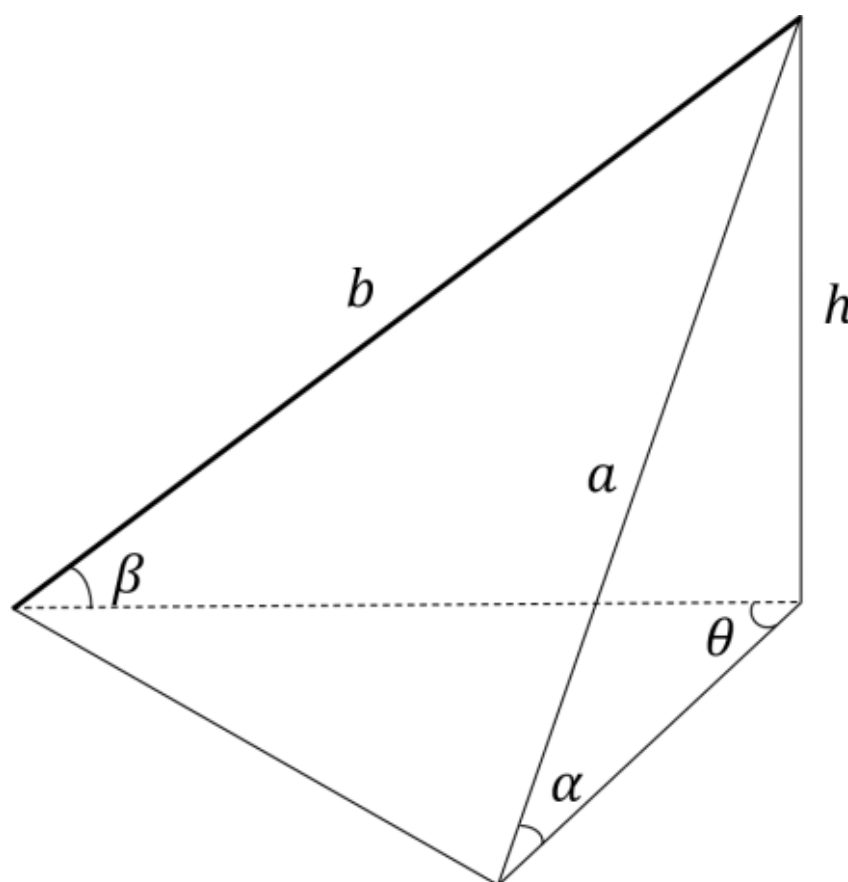


Figure 6. Schematic diagram of mountain terrain variables: a represents the straight line along the slope gradient of the mountain, b represents the road, h represents the height of the indicated position, θ represents the road-gradient angle, α represents the slope at the steepest direction directly obtained from a DEM and β represents the required road-direction slope.

4. Research Region and Experimental Environment

4.1. Research Region

Jiuzhaigou county belongs to the Tibetan Qiang Autonomous Prefecture of Ngawa in Sichuan province, China. It is located on the eastern edge of the Qinghai–Tibet Plateau and the northeastern part of Ngawa prefecture, covering a total area of 5286 km². It has a complex geological background with an extensive carbonate distribution, developed folds and fractures, strong neotectonics movement, large crustal uplift and a variety of complex forces creating a variety of landforms. The river valleys in Jiuzhaigou county are vertical and horizontal. The terrain is high in the northwest and low in the southeast; it is dominated by high mountains. The topography changes throughout this region and the elevation difference is up to 2000 meters. There are many lakes, waterfalls and calcified beach streams in the ditches and virgin forest covers more than half of the area.

A topographic map of Jiuzhaigou county is shown in Figure 7. The DEM data in the figure come from the hi-res terrain corrected data of ALOS PALSAR (<https://search.asf.alaska.edu/>). The region outline comes from the national catalogue service for geographic information of China (<http://www.webmap.cn/>). It can be seen that the terrain of Jiuzhaigou county is complex so that special processing is needed in road extraction.

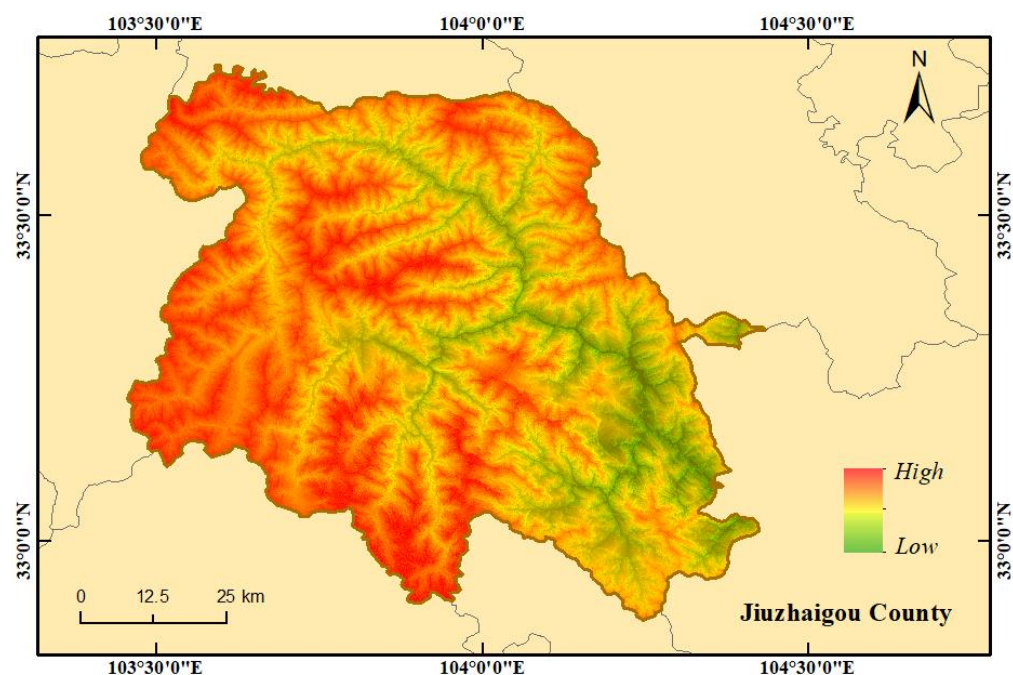


Figure 7. Topography of Jiuzhaigou county. We used a rendering method to display the DEM, so there is not a precise numerical representation, but rather high and low elevation.

4.2. Experimental Environment

We used the Windows 10 system with an NVIDIA GeForce RTX 2070 8G graphics card as the experimental environment. We used Python and PyTorch framework to implement the network models.

4.3. Experimental Using Samples Locations

We selected 90 sample points in the Jiuzhaigou county region to obtain sample images of 1000 pixels $\times$ 1000 pixels. After that, the labels were drawn manually according to the RGB images. Finally, 90 image samples and corresponding labels were obtained. In order to avoid instability caused by the limited amount of data, the labels were organized into two groups, each with a total of 90 images but the data used for training and testing is different. From each group, 60 images were used as training data and 30 images were used as test data. The test data used in the two groups were completely different. We compared the results of our method with those of the U-Net and D-LinkNet networks, which perform well in road extraction [37]; D-LinkNet achieved first place in CVPR2018: deepglobe road extraction challenge and U-Net is a classic network structure widely used for road information extraction [44]. When comparing the networks, the model effects before and after adding the CLAHE algorithm were also compared.

4.4. Experiment over a Complete Region

We conducted experiments using the complete high-resolution data of Jiuzhaigou county. Road extraction was performed on all images in Jiuzhaigou county region using moving windows. After extraction, the road extraction raster map of the complete region was obtained and the raster map was converted into a shapefile. Then we used a postprocessing operation with terrain constraints to improve road extraction.

5. Accuracy Evaluation Scheme

Since our study involved the extraction of complete road information in a mountainous region, a single accuracy evaluation scheme cannot reasonably evaluate the methods we proposed. We used three accuracy evaluation methods to evaluate the results of mountain road extraction. All of the evaluation methods have the same basic evaluation indicators.

5.1. Basic Accuracy Evaluation Indicators

The main evaluation indicators are precision, recall and comprehensive accuracy score. We took the sample label as the actual value and the output result of our methods as the predicted value. The precision and recall are obtained by:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

where *FP* indicates the number of extraction errors. *FN* indicates the number of correct values that have not been extracted. On this basis, the *F1* score can be obtained by:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Precision indicates whether the extracted road was extracted correctly, and recall indicates whether the real road was extracted completely. The *F1* score can comprehensively reflect the extraction accuracy, so we used the *F1* score to judge accuracy.

5.2. Specific Evaluation Method

5.2.1. Cross Validation Based on Raster Data

This method is a commonly used road extraction accuracy verification method that verifies every pixel of the samples, so it is convincing and credible. Since our study used fewer samples, in order to make the experimental results more stable, a cross-validation method was adopted. We used the samples introduced in Section 4.3 for this. All samples were divided into two groups, the test samples in one group are used as part of the training samples in the other group, and the training samples of the two groups are different from each other. We tested the two groups of samples separately and calculated the average of the accuracy of the results.

5.2.2. Large-Scale Validation on Point Data

Although the method in Section 5.2.1 can perform pixel-by-pixel inspection, there will be some concentrated validation regions and some sparse validation regions. At the large regional scale, some areas may be missed. Point samples in mountainous regions can make the sample distribution more uniform so that the extraction accuracy can be expressed more completely. Therefore, 200 sample points were selected here and they were relatively evenly distributed over the study region. These points were verified on Google Earth and field survey data of Jiuzhaigou.

5.2.3. Validation Using OSM Data

We used OSM data to test the road extraction results and improve the reliability of the accuracy measurement. OSM is a famous world map that can be used freely under an open license agreement. As shown in Figure 8, the OSM data were consistent with the shape and position of the actual road, but there were some deviations in the specific spatial position. To ensure accurate verification, we calculated a 100 m buffer of the OSM data and used the buffer of OSM and the road extraction results to calculate the accuracy value. We calculated the length of the OSM data and the length of the matching extraction result, then calculated the accuracy by comparing their lengths.



Figure 8. Diagram of OpenStreetMap (OSM) data.

6. Results and Discussion

6.1. Results of the Raster Samples

The experimental results for the raster samples are shown in Table 1. All experiments used pretrained models. According to the characteristics of the structure, we used the pretrained model of VGGNet [45] in U-Net, and the pretrained model of ResNet in other networks.

Table 1. Accuracy of different experimental models.

Pretreatment	Network	Group	Precision	Recall	F1	mF1
-	U-Net	1	0.8501	0.8354	0.8427	0.8145
		2	0.7879	0.7845	0.7862	
-	D-LinkNet	1	0.8966	0.8305	0.8622	0.8529
		2	0.8453	0.8418	0.8436	
HE	D-LinkNet	1	0.8559	0.8604	0.8582	0.8413
		2	0.8079	0.8416	0.8244	
CLAHE	D-LinkNet	1	0.8969	0.8338	0.8642	0.8579
		2	0.8536	0.8494	0.8515	
CLAHE	DSDNet	1	0.8979	0.8463	0.8713	0.8631
		2	0.8567	0.8531	0.8549	

It can be seen from Table 1 that the extraction accuracy of the D-LinkNet network was higher than the U-Net series. The accuracy of U-Net in the two groups was not stable, while the accuracy of extraction in D-LinkNet is relatively stable. After adding the histogram equalization method, the extraction accuracy of the D-LinkNet network reduced. Due to the influence of mountain shadows and poor image quality, the globally enhanced histogram equalization method does not have a positive effect. However, after using the CLAHE algorithm, the accuracy greatly improved, which shows the effectiveness of the CLAHE algorithm. The accuracy of road was highest when using the DSDNet proposed in this paper as the backbone network, showing that feature extraction benefited from the SE-module and dense connection upsampling method.

Figure 9 shows the details of the extraction effect from different methods. In Figure 9a, the results of multiple network structure extraction showed road interruption, but it did not appear in DSDNet (CLAHE), and the result of DSDNet (CLAHE) was the closest to the ground truth. In (b), it can be seen that there is some interference information beside the road, which is non-road based on a manual judgment. In the results of the neural network extraction, there are different degrees of mis-extraction. The result extracted by DSDNet (CLAHE) had the least false detection and was the closest to ground truth. In (c), there is a small road next to the main road. The path is narrow with many trees covering it, so it is difficult to identify. Each network structure had different degrees of missed road detection, but among them D-LinkNet (HE) and DSDNet (CLAHE) were most effective. In (d), the spectral characteristics of the road and the background are very similar. There

are obvious misdetections by the U-Net and D-LinkNet networks. After using histogram equalization processing, the effect was significantly improved. The CLAHE method had a greater improvement in the extraction effect.

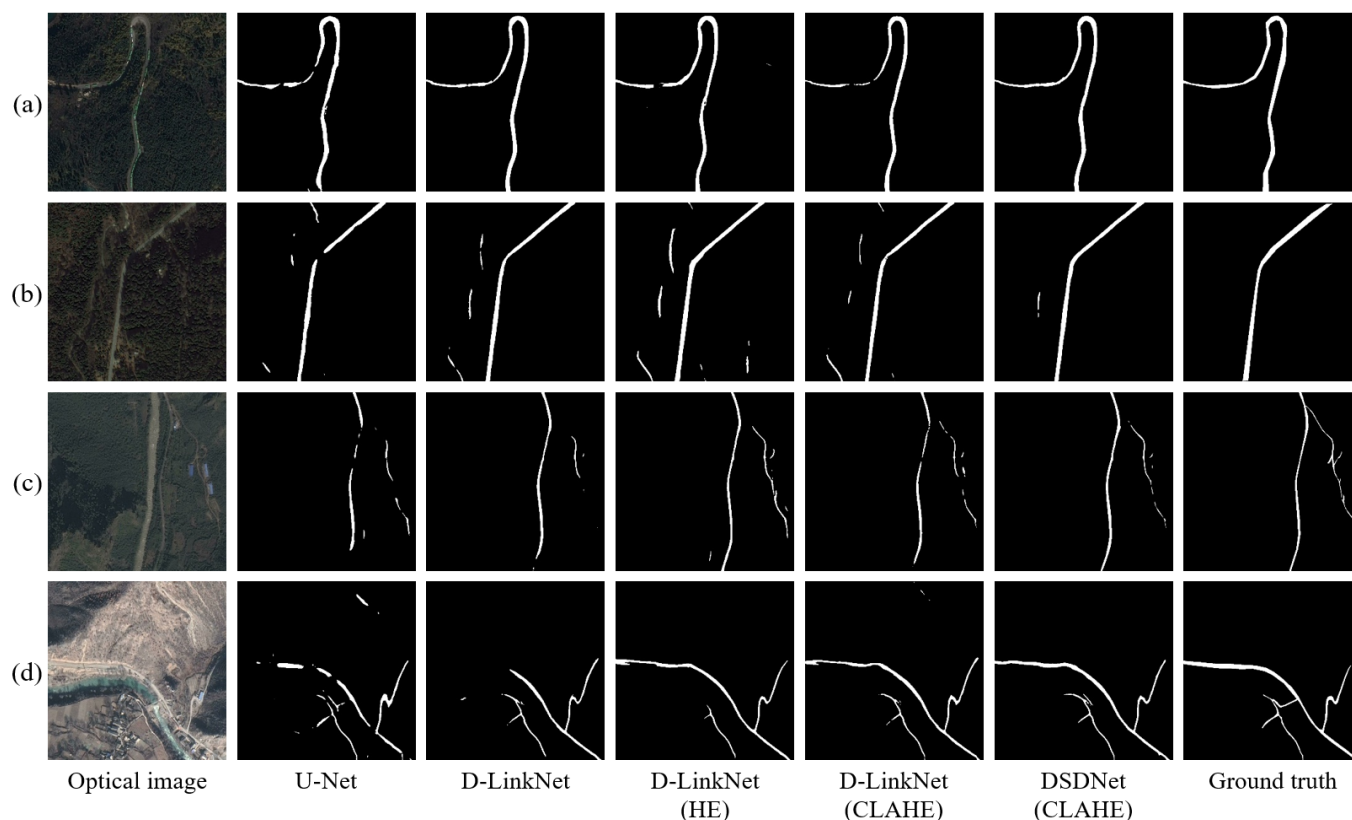


Figure 9. Diagram of the results from experimental samples: (a–d) shows the extraction of different samples.

It can be seen from these samples that the method proposed in this paper can extract roads very well. Moreover, compared with manually drawn roads, the road extraction contour obtained by the neural network is smoother and more in line with the actual situation.

6.2. Results for the Complete Region

Although we selected raster samples as evenly as possible, it is still inevitable that some information will be missed, especially with a complex mountainous background. Therefore, the road extraction experiment was carried out over the complete region of Jiuzhaigou county, and a complete road extraction map was obtained. We used 200 verification points (Figure 10) and the results of verification are shown in Table 2. For the reason that most parts of the mountain areas are not roads, uniform or random points cannot fully reflect the accuracy, so the selected verification points were not uniform, nor were they randomly generated. However, these points were relatively evenly distributed throughout the research region and had undergone strict authentic certification.

According to Table 2, DSDNet (CLAHE) was better than the D-LinkNet network in terms of overall F1 accuracy, and its advantage was mainly reflected in the recall rate. Some mountainous terrain was also extracted by mistake because of its similarity to roads. After using terrain constraints, the precision was significantly improved, which shows that terrain constraints could effectively reduce false detections. Although recall was slightly reduced after using terrain constraints, its extraction accuracy significantly improved and was 5.68% and 3.44% higher than D-LinkNet and DSDNet (CLAHE) without postprocessing, respectively.

Figure 11 shows the details of road extraction by different methods. In the first row, some parts of the road were obscured by mountain shadows, resulting in a significant spectral difference. In the algorithm with no CLAHE enhancement, the shadow areas had obvious missed detections. CLAHE-preprocessing substantially improved road recognition rates. In the second and third rows, there were no roads, but some areas were incorrectly extracted as roads in the method without terrain constraints. Many mountainous features were very similar to roads in the mountainous region. After using terrain constraint postprocessing, false detection was reduced.

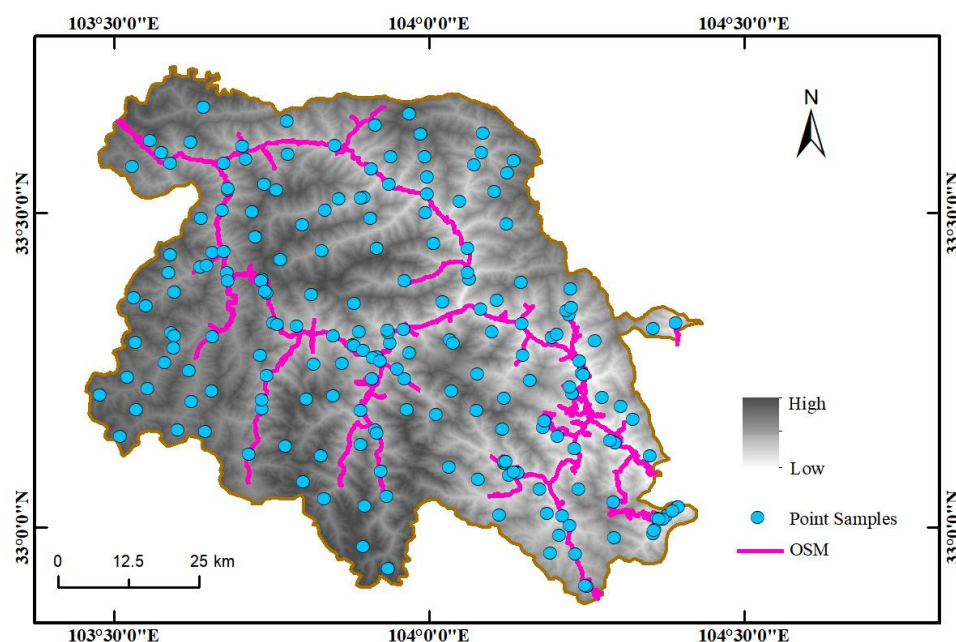


Figure 10. Diagram of point samples and OSM roads. We used a rendering method to display the DEM, so there is not a precise numerical representation, but rather high and low elevation.

Table 2. Accuracy of different experimental models.

Method	Post-Processing	Precision	Recall	F1
D-LinkNet	-	0.7981	0.8218	0.8098
DSDNet (CLAHE)	-	0.7647	0.9010	0.8273
DSDNet (CLAHE)	Terrain Constraints	0.8318	0.8812	0.8558

We used OSM data to further test the accuracy of road extraction. Since OSM data in mountainous regions was not very detailed, we tested the extraction accuracy of the main road network (Figure 9) and we only calculated the recall. The verification results are shown in Table 3. It can be seen that the extraction results using DSDNet (CLAHE) were acceptable and higher than the result of D-LinkNet. Therefore, we could conclude that the method we proposed had a good practical application effect.

Figure 12 shows the results (polylines) from the different methods for a typical area of the region. The figure shows the road extraction accuracy using our proposed method in a shaded area. In the shaded area, the method we proposed performed better than the original D-LinkNet method. As the OSM data are public data and not determined by the experimenter, the validation of the OSM data further verified the effectiveness of the method we proposed.

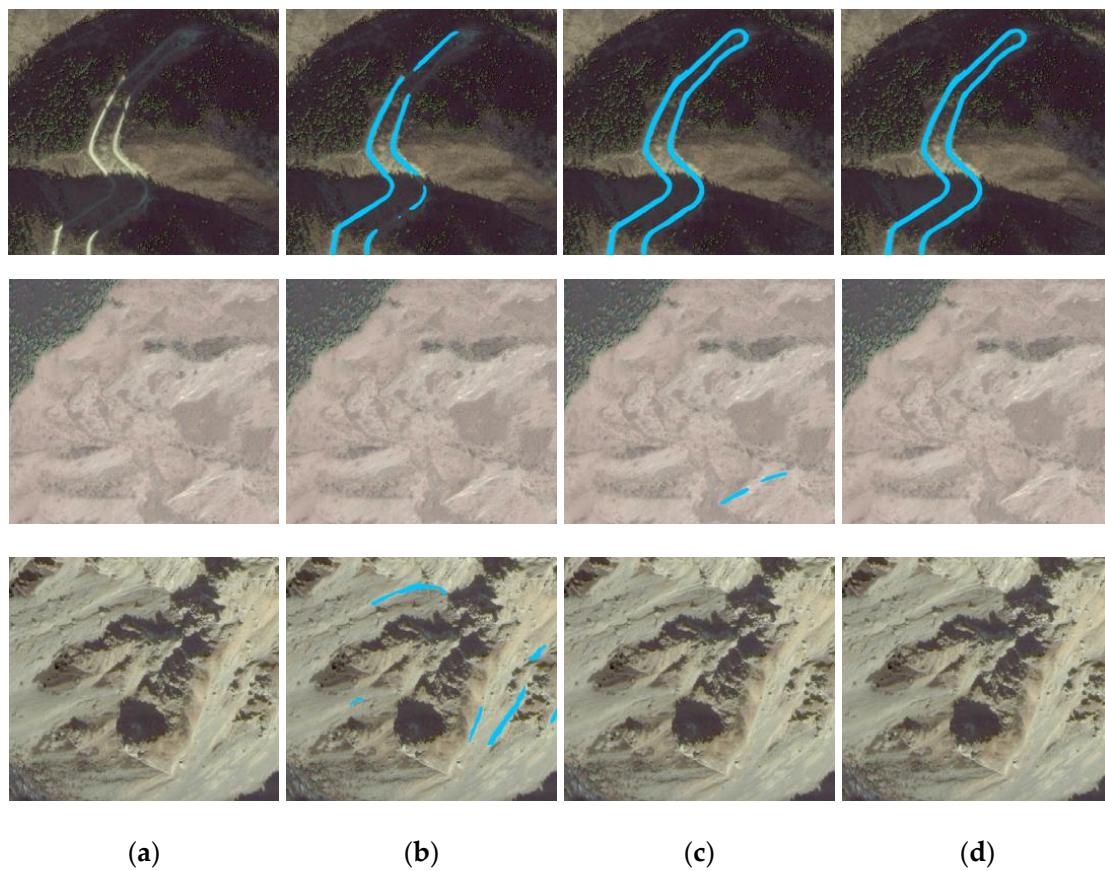


Figure 11. Diagram of the results for the complete region: (a) Optional image; (b) result of D-LinkNet; (c) result of DSDNet with CLAHE method—the image has not been processed with terrain constraints and (d) result of DSDNet with CLAHE method—the image has been processed with terrain constraints.

Table 3. Accuracy of different experimental models.

Method	Post-Processing	Recall
D-LinkNet	-	0.8673
DSDNet (CLAHE)	-	0.8854
DSDNet (CLAHE)	Terrain Constraints	0.8801

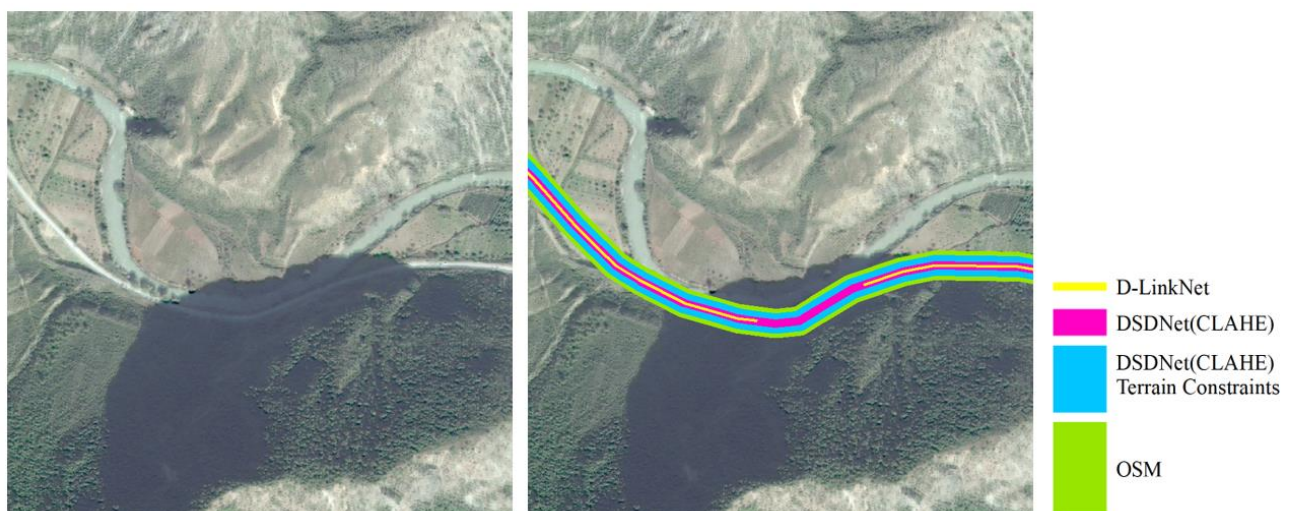


Figure 12. Comparison of results for each method used in a typical road area.

Synthesizing multiple evaluation schemes, the extraction effect of mountain roads proposed in this paper had the highest extraction accuracy and a high practical value.

7. Conclusions and Future Lines of Research

According to the experiments we conducted, the method proposed in this paper was able to accurately extract road from remote sensing data in a mountainous region. The CLAHE algorithm improved results especially for the quality degradation caused by clouds and fog and the interference of road spectrum information caused by mountain shadows. Through DSDNet, road features can be more accurate. By using postprocessing with terrain constraints, the problem of false detection can be reduced well.

Figure 13a,b show that our method had good results in the case of tree shadows and vehicle interference. However, Figure 13c shows that some tall trees could completely obscure longer roads (sometimes more than 1 km), which are difficult to extract even with visual interpretation. We considered combining multisource (such as SAR data) and multiperiod data for comprehensive judgment and extraction under these circumstances. In addition, the postprocessing in this article was limited by the threshold method, which requires multiple experiments to reach the optimal values for thresholds. In the future, we will consider studying general methods to determine thresholds or use shallow machine learning algorithms (such as SVM) to replace the threshold method.

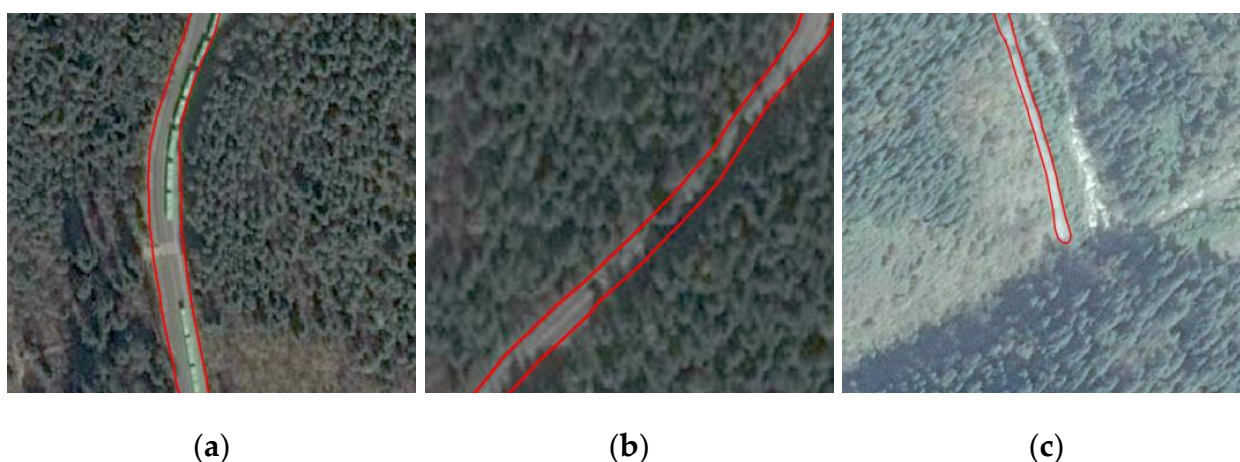


Figure 13. Typical details of road extraction using our proposed model: (a) Extraction result of road with vehicles; (b) extraction result of road with tree shadow; (c) extraction result of road in the area with tall trees.

Author Contributions: Z.X. completed the main algorithm design; Z.S. completed the optimization of the algorithm; Z.X. and Y.L. (Yang Li) implements the algorithms; L.X. preprocessed the data; Z.X., H.W. and S.L. conducted some of the experiments on road detection; Z.X., S.J. and Y.L. (Yating Lei). wrote and edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by National Key Research and Development Project of China (2017YFB0504204, 2018YFB0505000), National Natural Science Foundation of China (41971375), Xinjiang Uygur Autonomous Region flexible talent award in 2018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were used in this study. The OSM data can be found at: [<https://www.openstreetmap.org/>]. The DEM data can be found at [<https://search.asf.alaska.edu/>]. The region outline of Jiuzhaigou county can be found at [<http://www.webmap.cn/>].

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

Abbreviations

The following abbreviations are used in this manuscript:

VHR	Very High Resolution image
OSM	OpenStreetMap
HE	Histogram Equalization
AHE	Adaptive Histogram Equalization
CLAHE	Contrast Limited Adaptive Histogram Equalization
SVM	Support Vector Machine
DEM	Digital Elevation Model

References

1. Zhou, M.T.; Sui, H.G.; Chen, S.X.; Wang, J.D.; Chen, X. BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 288–306. [\[CrossRef\]](#)
2. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sens.* **2020**, *12*, 1444. [\[CrossRef\]](#)
3. Lian, R.B.; Wang, W.X.; Mustafa, N.; Huang, L.Q. Road extraction methods in high-resolution remote sensing images: A comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5489–5507. [\[CrossRef\]](#)
4. Lu, J.W.; Hu, J.L.; Zhou, J. Deep metric learning for visual understanding: An overview of recent advances. *IEEE Signal Process. Mag.* **2017**, *34*, 76–84. [\[CrossRef\]](#)
5. Fu, K.; Peng, J.S.; He, Q.W.; Zhang, H.X. Single image 3D object reconstruction based on deep learning: A review. *Multimed. Appl.* **2020**. [\[CrossRef\]](#)
6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [\[CrossRef\]](#)
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
8. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37. [\[CrossRef\]](#)
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
10. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241. [\[CrossRef\]](#)
12. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [\[CrossRef\]](#)
13. Zhang, X.; Han, X.; Li, C.; Tang, X.; Zhou, H.; Jiao, L. Aerial image road extraction based on an improved generative adversarial network. *Remote Sens.* **2019**, *11*, 930. [\[CrossRef\]](#)
14. Zhang, X.; Zhang, C.K.; Li, H.M.; Luo, Z. A road extraction method based on high resolution remote sensing image. In Proceedings of the 2020 International Conference on Geomatics in the Big Data Era (ICGBD), Guilin, China, 15–10 November 2020; pp. 671–676. [\[CrossRef\]](#)
15. Soni, P.K.; Rajpal, N.; Mehta, R. Semiautomatic road extraction framework based on shape features and ls-svm from high-resolution images. *J. Indian Soc. Remote Sens.* **2020**, *48*, 513–524. [\[CrossRef\]](#)
16. Gruen, A.; Li, H.H. Road extraction from aerial and satellite images by dynamic programming. *ISPRS J. Photogramm. Remote Sens.* **1995**, *50*, 11–20. [\[CrossRef\]](#)
17. Hinz, A.; Baumgartner, A. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 83–98. [\[CrossRef\]](#)
18. Zhao, Q.H.; Wu, Y.; Wang, H.; Li, Y. Road extraction from remote sensing image based on marked point process with local structure constraint. *Chin. J. Sci. Instrum.* **2020**, *41*, 185–195. [\[CrossRef\]](#)
19. Peng, B.; Xu, A.; Li, H.; Han, Y. Road extraction based on object-oriented from high-resolution remote sensing images. In Proceedings of the 2011 International Symposium on Image and Data Fusion, Tengchong, China, 9–11 August 2011; pp. 1–4. [\[CrossRef\]](#)
20. Yang, X.F.; Li, X.T.; Ye, Y.M.; Lau, P.Y.K.; Zhang, X.F.; Huang, X.H. Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 185–195. [\[CrossRef\]](#)
21. He, X.H.; Li, D.S.; Li, P.L.; Hu, S.K.; Chen, M.Y.; Tian, Z.H.; Zhou, G.S. Road extraction from high resolution remote sensing images based on EDRNet model. *Comput. Eng.* **2020**, 1–11. [\[CrossRef\]](#)

22. Bastani, F.; He, S.; Abbar, S.; Alizadeh, M.; Balakrishnan, H.; Chawla, S.; Madden, S.; DeWitt, D. Roadtracer: Automatic extraction of road networks from aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 19–21 June 2018; pp. 4720–4728. [\[CrossRef\]](#)
23. Qi, X.; Li, K.; Liu, P.; Zhou, X.; Sun, M. Deep attention and multi-scale networks for accurate remote sensing image segmentation. *IEEE Access*. **2020**, *8*, 146627–146639. [\[CrossRef\]](#)
24. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
25. Zhang, Y.; Xiong, Z.Y.; Zang, Y.; Wang, C.; Li, J.; Li, X. Topology-aware road network extraction via multi-supervised generative adversarial networks. *Remote Sens.* **2019**, *11*, 1017. [\[CrossRef\]](#)
26. Costea, D.; Marcu, A.; Slusanschi, E.; Leordeanu, M. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2100–2109. [\[CrossRef\]](#)
27. Tan, C.; Sun, F.C.; Kong, T.; Zhang, W.C.; Yang, C.; Liu, C.F. A survey on deep transfer learning. In Proceedings of the 2018 Artificial Neural Networks and Machine Learning (ICANN), Rhodes, Greece, 4–7 October 2018; pp. 192–196. [\[CrossRef\]](#)
28. Li, W.; Wu, G.D.; Du, Q. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2019**, *57*, 2305–2323. [\[CrossRef\]](#)
29. Diane, C.; Kyle, D.F.; Narayanan, C.K. Transfer learning for activity recognition: A survey. *Know. Inf. Syst.* **2014**, *57*, 537–556. [\[CrossRef\]](#)
30. Senthilnath, J.; Varia, N.; Dokania, A.; Anand, G.; Benediktsson, J.A. Deep TEC: Deep transfer learning with ensemble classifier for road extraction from UAV image. *Remote Sens.* **2020**, *12*, 245. [\[CrossRef\]](#)
31. He, H.; Yang, D.F.; Wang, S.C.; Wang, S.Y.; Liu, X. Road segmentation of cross-modal remote sensing images using deep segmentation network and transfer learning. *Ind. Robot.* **2019**, *46*, 384–390. [\[CrossRef\]](#)
32. Zhang, Q.; Kong, Q.; Zhang, C.; You, S.; Wei, H.; Sun, R.; Li, L. A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network. *Eur. J. Remote Sens.* **2019**, *52*, 572–582. [\[CrossRef\]](#)
33. Henry, C.; Azimi, S.M.; Merkle, N. Road segmentation in SAR satellite images with deep fully convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2020**, *15*, 1867–1871. [\[CrossRef\]](#)
34. Martins, V.S.; Kaleita, A.L.; Gelder, B.K.; da Silveira, H.L.F.; Abe, C.A. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 56–73. [\[CrossRef\]](#)
35. Robinson, C.; Hou, L.; Malkin, K. Large scale high-resolution land cover mapping with multi-resolution data. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12718–12727. [\[CrossRef\]](#)
36. Salberg, A.B.; Trier, Ø.D.; Kampffmeyer, M. Large-scale mapping of small roads in lidar images using deep convolutional neural networks. In Proceedings of the Scandinavian Conference on Image Analysis, Tromsø, Norway, 12–14 June 2017; pp. 193–204. [\[CrossRef\]](#)
37. Courtial, A.; Ayedi, A.E.; Touya, G.; Zhang, X. Exploring the potential of deep learning segmentation for mountain roads generalisation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 338. [\[CrossRef\]](#)
38. Abdollahi, A.; Bakhtiari, H.R.R.; Nejad, M.P. Investigation of SVM and level set interactive methods for road extraction from google earth images. *J. Indian Soc. Remote Sens.* **2018**, *46*, 423–430. [\[CrossRef\]](#)
39. Zhou, L.; Zhang, C.; Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the 31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–196. [\[CrossRef\]](#)
40. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851. [\[CrossRef\]](#)
41. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [\[CrossRef\]](#)
42. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 18th IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460. [\[CrossRef\]](#)
43. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883. [\[CrossRef\]](#)
44. Spolti, A.; Guizilini, V.; Mendes, C.C.T.; Croce, M.; Geus, A.; Oliveira, H.C.; Backes, A.R.; Souza, J. Application of u-net and auto-encoder to the road/non-road classification of aerial imagery in urban environments. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), Valetta, Malta, 27–29 February 2020; pp. 607–614. [\[CrossRef\]](#)
45. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.