



# A Land Cover Background-Adaptive Framework for Large-Scale Road Extraction

Yu Li <sup>1</sup>, Hao Liang <sup>1</sup>, Guangmin Sun <sup>1,\*</sup>, Zifeng Yuan <sup>1</sup>, Yuanzhi Zhang <sup>2,3</sup> and Hongsheng Zhang <sup>4</sup>

<sup>1</sup> Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> Laboratory of Deep-Space Exploration, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup> Department of Geography, The University of Hong Kong, Hong Kong, China

\* Correspondence: gmsun@bjut.edu.cn

**Abstract:** Background: Road network data are crucial in various applications, such as emergency response, urban planning, and transportation management. The recent application of deep neural networks has significantly boosted the efficiency and accuracy of road network extraction based on remote sensing data. However, most existing methods for road extraction were designed at local or regional scales. Automatic extraction of large-scale road datasets from satellite images remains challenging due to the complex background around the roads, especially the complicated land cover types. To tackle this issue, this paper proposes a land cover background-adaptive framework for large-scale road extraction. Method: A large number of sample image blocks (6820) are selected from six different countries of a wide region as the dataset. OpenStreetMap (OSM) is automatically converted to the ground truth of networks, and Esri 2020 Land Cover Dataset is taken as the background land cover information. A fuzzy C-means clustering algorithm is first applied to cluster the sample images according to the proportion of certain land use types that obviously negatively affect road extraction performance. Then, the specific model is trained on the images clustered as abundant with that certain land use type, while a general model is trained based on the rest of the images. Finally, the road extraction results obtained by those general and specific modes are combined. Results: The dataset selection and algorithm implementation were conducted on the cloud-based geoinformation platform Google Earth Engine (GEE) and Google Colaboratory. Experimental results showed that the proposed framework achieved stronger adaptivity on large-scale road extraction in both visual and statistical analysis. The C-means clustering algorithm applied in this study outperformed other hard clustering algorithms. Significance: The promising potential of the proposed background-adaptive network was demonstrated in the automatic extraction of large-scale road networks from satellite images as well as other object detection tasks. This search demonstrated a new paradigm for the study of large-scale remote sensing applications based on deep neural networks.

**Keywords:** road extraction; land cover; deep learning models; unsupervised clustering



**Citation:** Li, Y.; Liang, H.; Sun, G.; Yuan, Z.; Zhang, Y.; Zhang, H. A Land Cover Background-Adaptive Framework for Large-Scale Road Extraction. *Remote Sens.* **2022**, *14*, 5114. <https://doi.org/10.3390/rs14205114>

Academic Editor: Mohammad Awrangjeb

Received: 15 August 2022

Accepted: 4 October 2022

Published: 13 October 2022

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



**Copyright:** © 2022 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

Road carriage accounts for a large proportion of the modern transportation industry. According to the Intergovernmental Panel on Climate Change (IPCC) report, total freight transport emissions will triple by 2050 compared to 2010 levels, and freight demand is closely related to GDP [1]. In order to analyze social, economic, and geographical conditions, high-quality road network data are required; these are widely used in urban planning, transportation navigation, risk management, etc. [2,3].

Remote sensing is a timely and cost-effective way to map the earth's surface [4–9]. With the developments of image processing and machine learning approaches [10,11], many studies have been conducted to reduce labor costs and improve the stability of road extraction from RSI. In 1976, Bajcsy and Tavakoli firstly proposed an ad-hoc approach to

automatically recognize roads from satellite images [12]. Pioneering road extraction models based on RSI utilized image processing techniques such as edge extraction, segmentation, pruning, and skeleton extraction [13]. More recent road extraction models were based on machine learning algorithms, where texture and spectral features were used as the input. A back propagation (BP) neural network was proposed for road detection in [14], where the gray-scale co-occurrence (GLCM) was adopted to compute pixel-wise texture parameters and generate a pre-classified road grid map. The texture parameters and spectral features were combined to effectively detect the road regions. In [15], the multi-class support vector machine was used for building and road detection. The integrated framework for urban road centerline extraction was proposed, which consists of spectral-spatial classification, local statistical analysis, linear kernel smoothing regression, and tensor voting [16]. Yin et al. proposed a direction-guided Ant Colony optimization method to intergrade both object and edge properties of the very high-resolution images for higher performance of road extraction [17]. These traditional methods based on hand-crafted feature extraction, machine learning, and statistical analysis are relatively easy to implement and computationally efficient. However, their performance has been limited due to the variety of road and non-road objects, complexity of backgrounds, occlusions, and shadows [18–20].

Deep learning models, including deep convolutional neural networks (DCNNs), have recently proved their promising potential in road detection based on RSI with their representation capability for complex feature relationships. Mnih and Hinton adopted a neural network model based on Restricted Boltzmann Machines (RBMs), where unsupervised pretraining and post-processing were used for road extraction [20]. The U-Net [21] is an end-to-end full convolutional network, consisting of encoder and decoder sub-networks. The encoder extracts the features using convolution and pooling layers, while the decoder achieves precise positioning using de-convolution layers. In Zhang et al. [22], the deep residual U-Net was proposed for road detection, where the U-Net and residual networks were combined. The LinkNet [23] introduces a direct connection between the encoder and decoder to recover lost information by pooling layers. The spatial information is directly connected to the decoder on top of the encoder-decoder pyramid structure. This increases segmentation accuracy without increasing computational complexity. The advanced variant of the LinkNet, D-LinkNet, was proposed based on high-resolution RSI in Zhou et al. [24]. D-LinkNet retained the encoder of the LinkNet as the backbone. The stacked dilated convolution layers in cascade and parallel are used to replace the pooling layers, maximizing the receptive field while retaining the central information.

Sentinel-2 remote sensing image dataset was used to extract large and medium-sized roads in [25]. A split depth-wise separable graph convolutional network was proposed for vegetation cover and road extraction [26]. A cascaded attention-enhanced architecture was proposed to extract boundary-refined roads from remote sensing images and obtained state-of-the-art results on the high-resolution Massachusetts dataset [27]. The performance of deep learning models highly depends on the quantity and quality of the dataset. However, the datasets used in the previous studies were mostly at local or regional scales. Even though transfer learning and weakly supervised learning techniques can be used to alleviate this requirement [28], accurate extraction of large-scale road data remains challenging [29].

The main challenges in large-scale road detection include the variety of road and background appearance, limited road network labels, and expensive parallel computing power. As demonstrated in [30,31], the type of land cover influences the road extraction performance of the DCNNs. According to the No Free Lunch Theorem (NFLT) in machine learning, adapting to all kinds of research regions is commonly difficult for a deep learning-based model. Therefore, it is theoretically reasonable to establish a classification model that is specifically applied to a particular problem. Inspired by this, this paper proposes a background-adaptive road extraction framework, where the DCNN is trained for the specific regions with certain land use types. In our previous work [31], the OpenStreetMap (OSM) was used to automatically generate reliable ground truth of networks, and Esri 2020 Land Cover Dataset was taken as the background land cover information. The recent

development of cloud-based computing platforms provides a flexible and economical option for establishing deep learning projects. In this study, the Google Earth Engine (GEE) and the machine learning platform Google Colaboratory were used to implement the proposed model.

The key scientific questions to be solved in this study are: (1) whether it is beneficial to train the deep learning-based road extraction models adaptively; (2) how to cluster the sample images according to their land cover types to achieve the best adaptive training performance; (3) what is the relationship between the performance of the deep learning-based road extraction algorithm and land cover type over the study area.

In order to answer the above problems, the remainder of this paper is structured as follows: First, we introduce the study area and dataset used in the experiment. Secondly, we describe our methods and experimental design in detail. Then, we present the experimental results, evaluation, and discussions. Finally, a conclusion is drawn, in which the above key scientific questions are answered.

## 2. Study Area and Dataset

### 2.1. Study Area

The study area in this paper is selected within the “One Belt and One Road” initiative region, which covers countries in Asia and Eastern Europe. Accurately obtaining changes in road networks contributes to the cooperation and development of countries in this region, aiding integration into a cohesive economic zone via infrastructure, cultural exchanges, and trade.

As shown in Figure 1, the data samples with various land cover types were collected from randomly selected countries, including Mongolia, Russia, Cambodia, Armenia, and China. The statistics of the proportions of all land cover types in the study area are shown in Figure 2, showing that trees, crops, scrub/shrub, and built area account for a majority of the proportion. Bare ground, snow/ice, and clouds accounted for a relatively very little proportion.

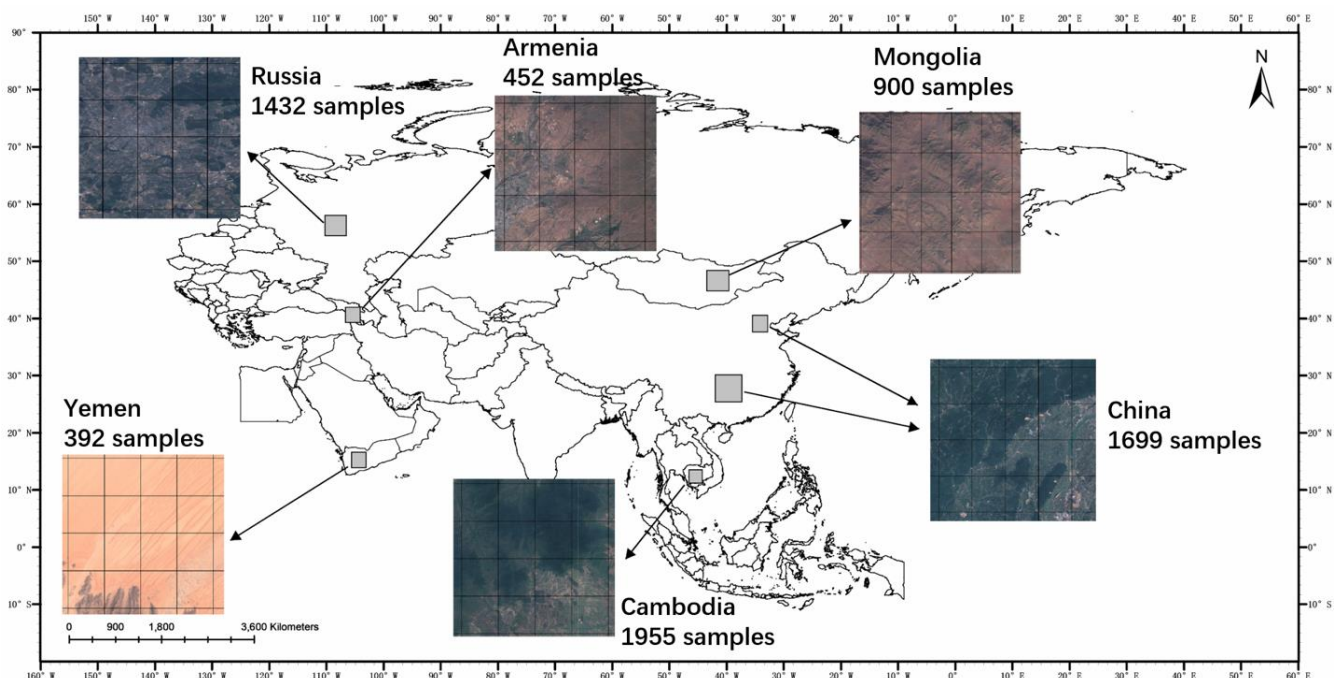
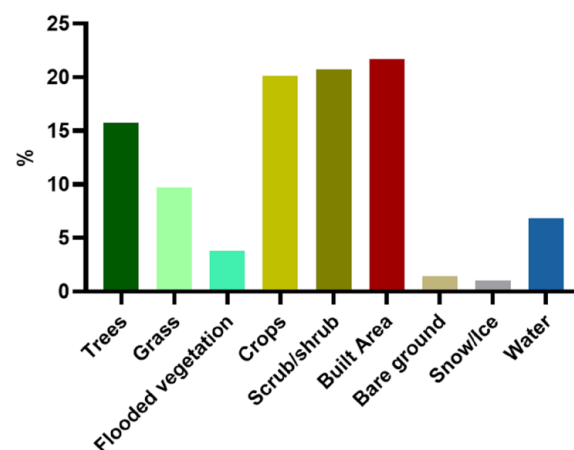


Figure 1. The study area and geographical distribution of samples.



**Figure 2.** The proportions of land cover types.

## 2.2. Dataset

All the images in the datasets were clipped with a grid of  $1024 \times 1024$  pixels, obtaining 6820 sample images in total. The datasets used in this study include Sentinel-2 image, Esri 2020 Land Cover, and OSM data, as shown in Table 1.

**Table 1.** Details of the datasets used in this study.

| Dataset                         | Data Time                           | Data Type                                                                                                                                                                     | Spatial Resolution (m) |
|---------------------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Sentinel-2                      | 1 January 2019–<br>31 December 2019 | Multispectral bands:<br>Blue (496.6 nm (S2A)/492.1 nm (S2B)),<br>Green (560 nm (S2A)/559 nm (S2B)),<br>Red (664.5 nm (S2A)/665 nm (S2B)),<br>NIR (835.1nm (S2A)/833 nm (S2B)) | 10                     |
| Esri 2020<br>Land Cover Dataset | 2020                                | Land use type classification map:<br>water, trees, grass, flooded vegetation, crops,<br>scrub/shrub, built area, bare ground,<br>snow/ice, and clouds                         | 10                     |
| OSM                             | 2019                                | Road data in vector format                                                                                                                                                    | -                      |

In this study, Sentinel-2 level-1C products provided by the European Space Agency were used to train the DCNN model. The (mostly) cloudless Sentinel-2 images of the study area were preprocessed using the Google Earth Engine (GEE) [32], where the visible bands (B2, B3, B4) with 10m resolution were used. Multiple images collected at different times for one year (1 January–1 December 2019) were median-filtered to obtain reliable images [33], under the assumption that the road regions are not changed. The mean filter allows for avoiding the over-saturation and underestimation of the vegetation due to phenological changes [34].

The OpenStreetMap (OSM) [35] is an open dataset collected by non-expert users who have a certain understanding of the region using various devices, such as satellite images, hand-held GPS devices, aerial photos, and satellite images. Although the OSM dataset is a large-scale dataset, it has inconsistent conditions for road data. According to [26], the OSM data is inconsistent, but this is, however, an acceptable amount for road extraction tasks. The original OSM data (vector form) is rasterized to be suitable for pixel-wise road extraction tasks. A buffer with a diameter of 10m was used considering the pixel resolution and sub-pixel extraction accuracy of Sentinel-2 data [35].

The global land cover dataset, Esri 2020 land cover dataset, was used as the prior knowledge of land cover types, which was derived from 10m Sentinel-2 data by using deep learning methods [35]. The dataset consists of ten land cover types (clouds, trees, water, built area, flooded vegetation, grass, snow/ice, crops, bare ground, and scrub/shrub) with

an overall accuracy of 85%. We assumed that the land cover type in 2020 had not changed much from 2019 when the Sentinel-2 images were acquired from a large-scale perspective.

### 3. Methods

#### 3.1. The Land Cover Background-Adaptive Framework

It has been proved that the proportion of certain background land use types in RSI affects road extraction performance [30]. Therefore, training a universal deep learning network is always difficult to obtain satisfactory road extraction performance in regions with a large variety of land cover backgrounds, especially for large-scale study sites. To tackle this, in this study, we propose a background-adaptive framework. A clustering algorithm based on Fuzzy C-means is proposed to cluster the image sample according to their main land cover type. Then, the proposed framework trains the deep learning models specifically for certain land use type that has a relatively large negative effect on road extraction performance. Finally, the specifically trained modes are used to process those image samples that are dominated by this land use type for better adaptivity.

As depicted in Figure 3, the proposed road extraction framework consists of three steps: preprocessing, land cover background clustering, model training, and model testing and evaluation. The following are the implementation details of the proposed study:

(a) Preprocessing:

The preprocessing includes data sample selection, registration, and clipping. The remote sensing images are mosaiced and clipped to a fixed size, and then the OSM data are rasterized and registered with the remote sensing images.

(b) Land cover background clustering:

The sample images are clustered by the proportion of certain land use types that obviously negatively affect road extraction performance. In this study, for instance, images mainly covered by Scrub/shrubs and Built areas are clustered out for training the specialized deep learning network. The clustering methods used in this study are detailed in the following Section 3.2.

(c) Model training:

Specific models are trained based on the images clustered as the certain land cover type, while a general model is trained based on the rest of the images. For this study, images that are mainly covered by Scrub/shrubs and Built areas are clustered to train the specific model-1 and specific model-2, respectively.

In this study, three typical deep learning-based semantic segmentation models are employed for road extraction, including U-Net, LinkNet, and D-LinkNet.

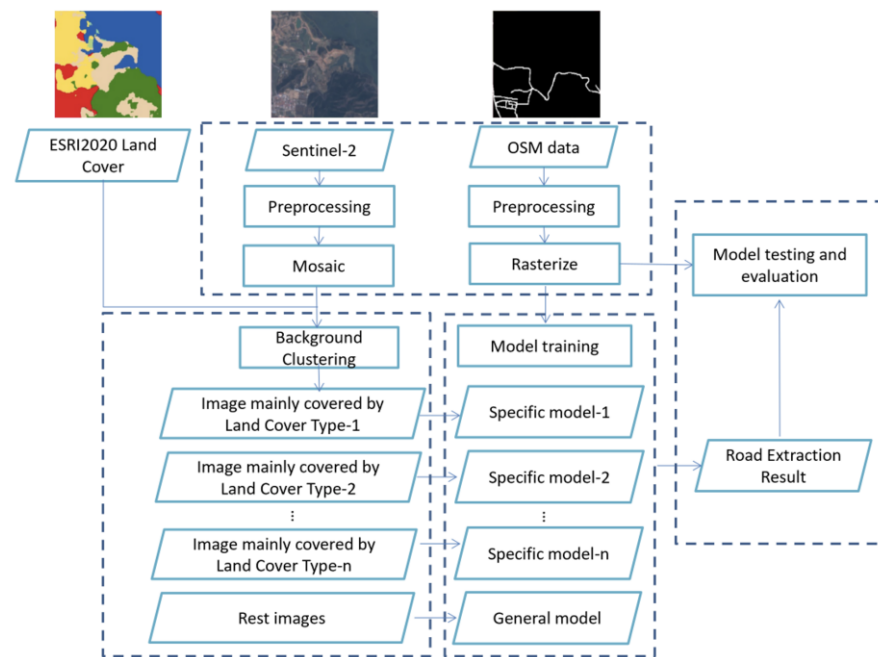
Adam optimizer [36] and the binary cross-entropy dice loss were used to train the road extract models. The cross-entropy dice loss is a combination of the binary cross-entropy loss  $L_{BCE}$  and dice loss  $L_{Dice}$ , as follows:

$$L = L(y, y_i)_{BCE} + L_{Dice} \quad (1)$$

$$L(y, y_i)_{BCE} = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (2)$$

$$L_{Dice} = 1 - \frac{2TP}{2TP + FP + FN} \quad (3)$$

where  $N$  represents the number of samples. The dice loss measures the imbalance between the road and background, using true positives ( $TP$ ), false positives ( $FP$ ), and false negatives ( $FN$ ).



**Figure 3.** The framework of the proposed background-adaptive framework for large-scale road extraction.

(d) Model testing and evaluation:

In the road extraction process, the remote sensing images are preprocessed and clustered similarly. Then, the specific models and the general model are implemented on them, respectively. Finally, the road extraction results are obtained by combining the output of different models. In this study, the road extraction results are evaluated by indexes introduced in the upcoming Section 3.3.

### 3.2. Land Cover Background Clustering

One of the keys of the proposed land cover background-adaptive framework is to identify the sample images that need to be trained and classified with the specific models. Fuzzy C-means (FCM) soft clustering is used to cluster image samples according to the main land cover types of their background [37,38]. Unlike hard clustering algorithms such as K-means [39], FCM generates flexible clusters suitable for cases where the objects are difficult to classify into distinct clusters.

FCM clustering assigns membership to each datum with respect to clusters according to the distance between the datum and the centers. The closer the datum is to a particular cluster center, the higher the membership it has to the cluster center. The objective is to minimize the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (4)$$

where  $m \geq 1$  is a real number named fuzzifier (in this paper,  $m = 2.5$ ). For the  $i$ th data  $x_i \in \mathbb{R}^d$ ,  $u_{ij}$  indicates the degree of membership of  $x_i$  in cluster  $j$ , of which the center is  $c_j \in \mathbb{R}^d$ .  $\|\cdot\|$  indicates any distance metric.

The Euclidean distance is used here to calculate the similarity between the center and the data:

$$D(x, y) = \left( \sum_{v=1}^n |x_v - y_v|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty \quad (5)$$

where  $p$  is the order parameter which is set as 2 in this experiment.

The objective function (1) is iteratively optimized to conduct the fuzzy partitioning, where  $u_{ij}$  and  $c_j$  are iteratively updated as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (6)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad (7)$$

In this study, the input feature  $x_i$  for clustering is the proportion of certain land cover types. Two widely used unsupervised clustering algorithms, K-means and DBSCAN, are also considered in this study to evaluate the advantage of the proposed FCM clustering method. K-means clustering is one of the most classic algorithms. It is efficient and easy to interpret, while it tends to reach local optimum and is sensitive to the outlier value [36]. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a representative density-based clustering algorithm. Compared with the K-means algorithm, it can better cope with noise and detect clusters with a variety of shapes. However, it is not suitable for the cases when the density distribution of the dataset is uneven and the difference between clusters is large [40].

### 3.3. Validation and Accuracy Assessment

Four commonly used evaluation indexes for semantic segmentation [41], mIoU, *Precision*, *Recall*, and F1 score were selected to evaluate the performance of the proposed method.

Intersection over Union (IoU) is defined as the ratio between the intersection and the union of the predicted results and ground truth. The mIoU indicates the average of the IoU values for different classes, defined as follows:

$$\text{mIoU} = \frac{1}{k} \sum_{j=1}^k \frac{p_{jj}}{p_{ij} + p_{ji} + p_{jj}}, i \neq j \quad (8)$$

where  $k$  indicates the number of classes. The classes  $j$ ,  $p_{ij}$ ,  $p_{ji}$ , and  $p_{jj}$ , represent *FP*, *FN*, and *TP*, respectively.

*Precision* is defined as the ratio of *TP* over the positive prediction results ( $TP + FP$ ), while *Recall* is defined as the ratio of *TP* over the truly predicted results ( $TP + TN$ ), which are formulated for the class  $j$  as follows:

$$\text{Precision} = \frac{p_{jj}}{p_{ij} + p_{jj}}, i \neq j \quad (9)$$

$$\text{Recall} = \frac{p_{jj}}{p_{ji} + p_{jj}}, i \neq j \quad (10)$$

F1 score measures the reduction of the class imbalance between road and background, which is the harmonic mean of *Precision* and *Recall* defined as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

F1 score ranges between 0 and 1.

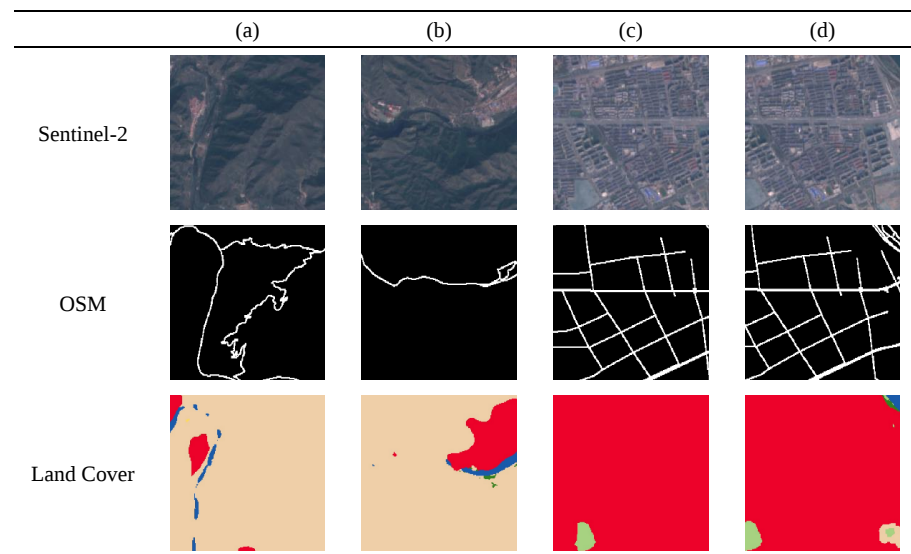
## 4. Results

### 4.1. Land Cover Background Clustering

The experiments were conducted on Google Colaboratory with the Google Earth Engine platform on a Linux system. An Nvidia Tesla T4 GPU was assigned, which has a

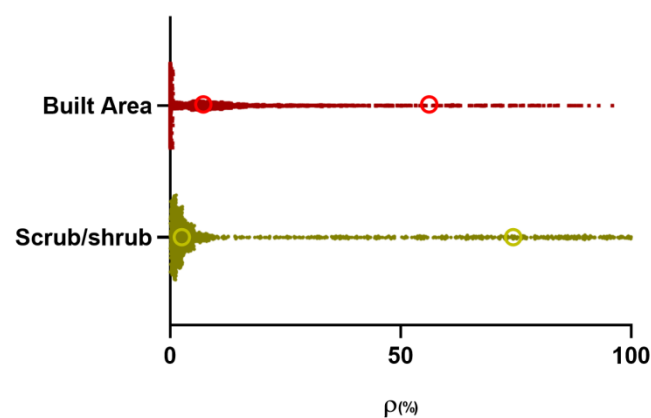
graphic memory of 16GB and 320 Turing Tensor Core and 2560 CUDA core and can achieve a floating-point computing power of approximately 65T.

Figure 4 shows examples of data samples from the study area, showing that denser roads are distributed in Built Area and are prone to be confused with other kinds of impervious surfaces. Moreover, the regions covered by Scrub/shrubs usually have less dense road distribution, which is likely to be confused with bare soil. In other words, those types of land cover significantly affect road extraction performance. Thus, in this study, the image samples with Scrub/shrubs and Built areas are clustered and used for training the specific model-1 and specific model-2, respectively.



**Figure 4.** Examples of regions covered by Scrub/shrubs (a,b) and Built Area (c,d).

Figure 5 shows the clustering results of FCM and the distribution of Scrub/shrubs and Built Area in the samples. The two circles on each scatter plot represent the two clustering centers derived by FCM. The clustering centers of Built Area were 56.519% and 6.126%, and the clustering centers of Scrub/shrubs were 74.629% and 4.333%. Therefore, the 1232 sample images of the high Scrub/shrubs proportion and 1027 samples of the high Built Area proportion were obtained.



**Figure 5.** Distribution and clustering centers (marked with cycles) of built-up area ratio and shrub/shrub ratio in the samples. Y-axis represents the Built Area and Scrub/shrub samples of land cover types, and X-axis represents the proportion of corresponding land cover types in the samples.

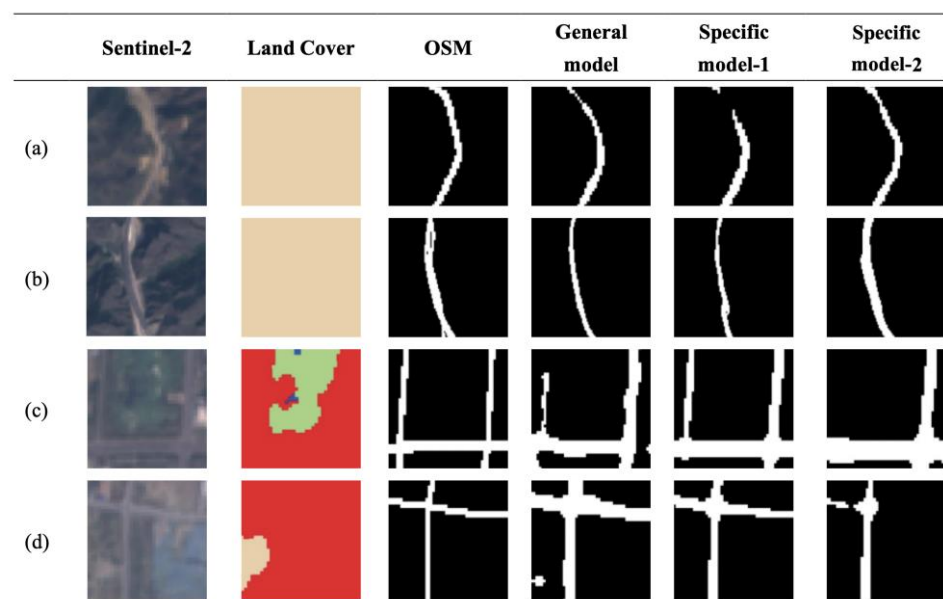
In this study, the proposed background-adaptive framework was applied based on three clustering algorithms: K-means, DBSCAN, and FCM. D-LinkNet, which achieved the

best road extraction performance, was used to train the road extraction models. As summarized in Table 2, the FCM obtained the best road extraction results, proving the advantage of the proposed clustering algorithm. Moreover, the processing time for clustering for the FCM was 1.607 s, which is negligible compared with the time consumed in training the deep learning-based algorithms.

**Table 2.** Road extraction results obtained by different clustering methods.

|                           | K-Means | DBSCAN | FCM           |
|---------------------------|---------|--------|---------------|
| <b>mIoU</b>               | 0.5533  | 0.5547 | <b>0.5694</b> |
| <b>Precision</b>          | 0.2945  | 0.3287 | <b>0.3467</b> |
| <b>Recall</b>             | 0.1867  | 0.3456 | <b>0.4646</b> |
| <b>F1 Score</b>           | 0.2285  | 0.3369 | <b>0.3601</b> |
| <b>Processing Time(s)</b> | 0.0364  | 3.5536 | <b>1.6075</b> |

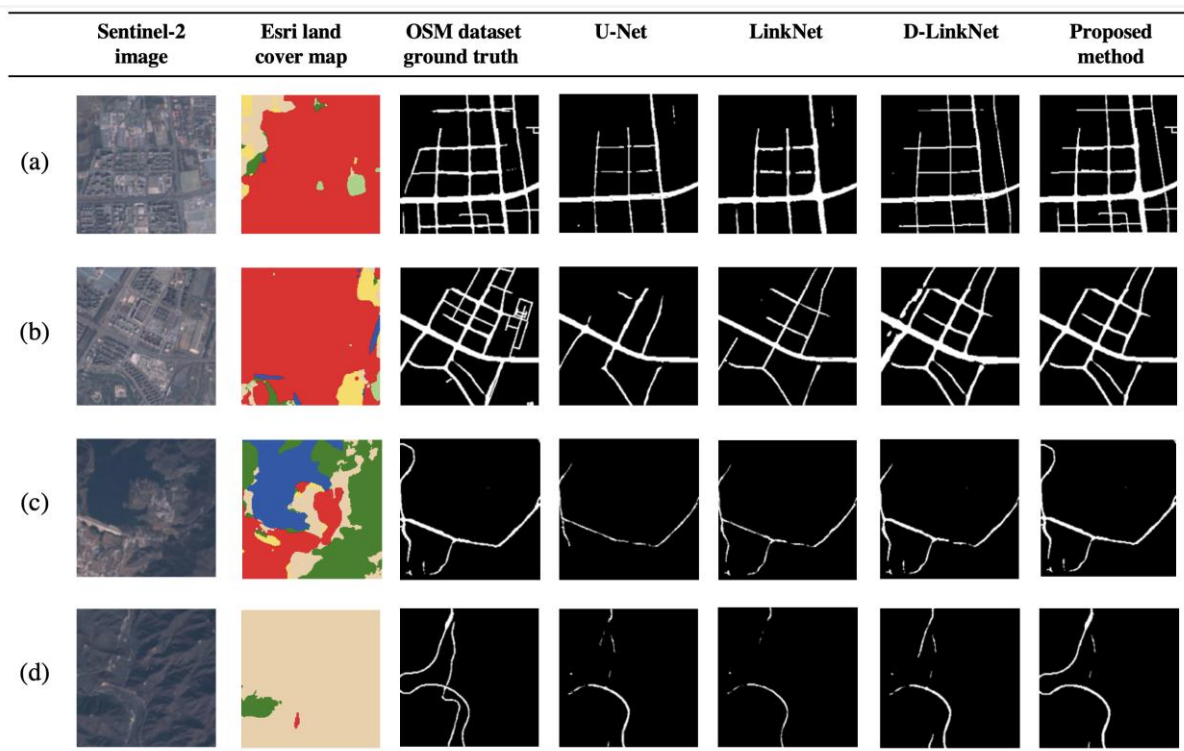
Figure 6 shows qualitative results obtained by different sub-models, showing that the Specific model-2 contributes the most to road extraction on sample images with high shrub proportion, while the Specific model-1 is optimal on samples with high Built Area proportion, which have a much lower missing rate and false alarm rate. These results show that the proposed framework works well for challenging samples.



**Figure 6.** Qualitative comparisons of road extraction by different sub-modules. (a,b) Scrub/shrubs; (c,d) Built Area.

#### 4.2. Large-Scale Road Extraction Results

Figure 7 presents road extraction results for large-scale road datasets. The proposed method provides a better segmentation result than D-LinkNet. Specifically, for regions covered by Built Area (colored red in Figure 7b), the proposed method has a much lower false alarm rate, while for regions covered by Scrub/Shrub (colored beige in Figure 7d), the lower missing rate is obtained.

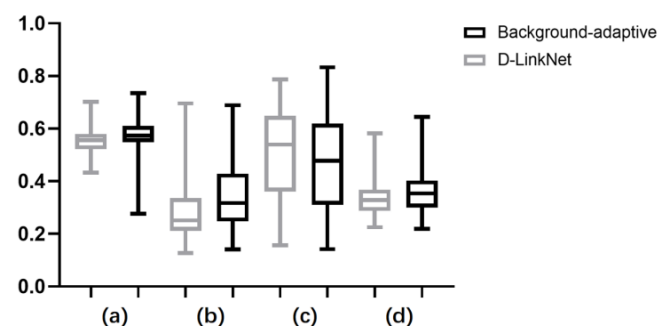


**Figure 7.** An example of the Road extraction results derived from different road extraction methods. (a,b) are two cases where Built Area is dominant; (c) is a case with very complicated land cover background, while (d) is a case where Scrub/Shrub is dominant.

Table 3 summarizes the road extraction performance in terms of objective evaluation indexes, where the proposed method obtained the highest mIoU, *Precision*, and F1 score. The statistical analysis of the box diagram (Figure 8) further confirmed the overall improvement of the proposed method towards the D-LinkNet. As can be seen in Table 3, the proposed background-adaptive road extraction framework has comparable inferring time with D-linkNet and LinkNet, with an average 1.3160 s per image of  $1024 \times 1024$  pixels.

**Table 3.** Comparisons of the road extraction performance derived by each method.

|                         | U-Net  | LinkNet | D-LinkNet | Background-Adaptive |
|-------------------------|--------|---------|-----------|---------------------|
| <b>mIoU</b>             | 0.5103 | 0.3098  | 0.5520    | 0.5694              |
| <b><i>Precision</i></b> | 0.3098 | 0.2927  | 0.2850    | 0.3467              |
| <b><i>Recall</i></b>    | 0.3054 | 0.3776  | 0.5076    | 0.4646              |
| <b><i>F1 Score</i></b>  | 0.2194 | 0.3008  | 0.3357    | 0.3601              |
| <b>Time (s)</b>         | 0.5695 | 1.0893  | 1.1352    | 1.3160              |



**Figure 8.** The box diagram for statistical analysis. (a) mIoU; (b) *Precision*; (c) *Recall*; (d) *F1 score*.

## 5. Discussion

### 5.1. Effects of Land Cover Types on Road Extraction Performance

It has been proved by previous experiments that land cover types have a significant effect on deep learning-based road extraction algorithms. To better verify the basis of the proposed method, this section conducts statistical analysis to evaluate the influence of the land cover types on road extraction performance. The proportion of each land type  $\rho_X$  is computed as follows:

$$\rho_X = \frac{N_X}{N_{all}} \quad (12)$$

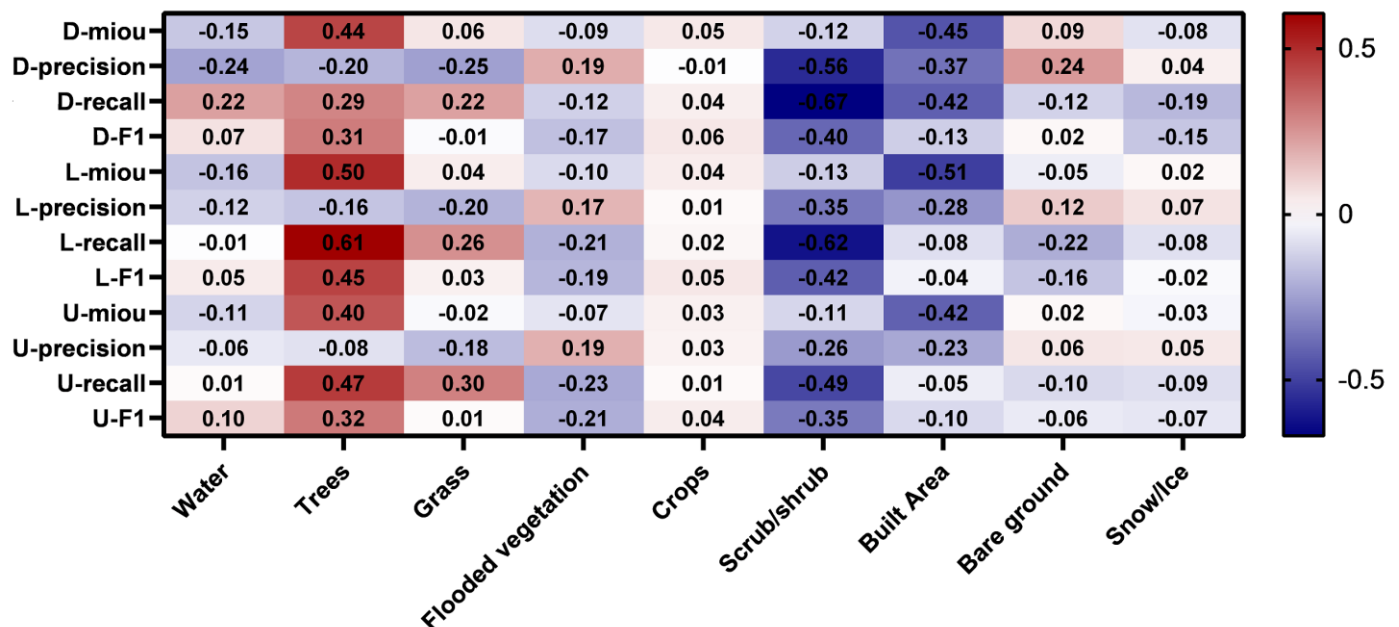
where,  $N_X$  and  $N_{all}$  are the number of pixels for the type  $X$  and the total number of pixels, respectively.

The Pearson Correlation Coefficient (PCC) [42] is used to analyze the relationship between  $\rho_X$  and the road extraction results, which is defined as follows:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} \quad (13)$$

where  $\bar{X}$  and  $\bar{Y}$  represent the means of  $X$  and  $Y$ , respectively. The value of PCC close to +1 or −1 indicates a higher positive or negative correlation between the two variables.

The analysis of PCCs between the land type and road extraction is given in Figure 9, where three models provide a similar trend. The proportions of Scrub/shrubs and Built Areas have a significantly larger negative correlation with the road extraction performance, while the proportion of Trees in the samples is more positively correlated with the road extraction performance. For the other land use types, the correlation is not obvious or inconsistent for different models. The results prove the advantage of the proposed method that adopts specifically trained models for challenging types.



**Figure 9.** The PCCs between the land cover types and the road extraction performance indexes. D-, L-, and U- indicate the results obtained by D-LinkNet, LinkNet, and U-Net, respectively.

### 5.2. Advantages and Limitations of the Proposed Framework

The key to the proposed framework is the appropriate clustering of images according to their land use background. The selection of specific land use types may differ in other research areas. The FCM used in the proposed method improves road extraction due to

its soft clustering characteristics. Another advantage of FCM is that it is an unsupervised method that does not need any predefined thresholds. The proposed background-adaptive framework outperforms the baseline network, D-LinkNet, in terms of all evaluation indexes except for Recall. This is because the D-LinkNet is trained to be more sensitive and tends to have a much higher false alarm. The OSM data was mainly derived by voluntary labeling, which mainly includes main roads. Nevertheless, the sensitivity of the proposed model can also be adjusted to meet the requirements of different applications. The proposed model will be fine-tuned and verified on more datasets in the future. It is worth noting that the influence of the land type on road extraction accuracy is related to sample size and the distribution of land cover type in the study area. In this experiment, samples were cropped by a grid of  $1024 \times 1024$  pixels. Theoretically, the denser the grid cutting, the purer the proportion of different land cover types in the samples. In future studies, we will explore the influence of different sample sizes on road extraction accuracy to determine the optimal sample size for large-scale road network extraction. Recently, there have also been several public road datasets with very high resolutions, such as the Massachusetts dataset and DeepGlobe dataset [25]. We will also verify the proposed method on more datasets in the future.

## 6. Conclusions

This paper proposes a novel framework for large-scale road extraction. With the help of increasing parallel computing capacities and the improvement of model efficiency provided by the cloud-based geospatial data processing platform Google Earth Engine and the machine learning platform Google Colaboratory, the application of deep learning-based methods has become more practical than ever before. The key innovations and contributions of this study are summarized as follows:

1. An obvious negative correlation between the proportion of Scrub/shrubs and Built Area and the road extraction accuracy is quantitatively discovered for the one belt and road region.
2. The Fuzzy C-means clustering algorithm is proven to achieve better land cover background clustering results than other hard clustering algorithms.
3. The proposed land cover background adaptive model achieves better road extraction results than compared models on large-scale road extraction tasks, obtaining improvements in the mIoU index by 0.0174, precision by 0.0617, and *F1* score by 0.0244.
4. The efficiency of the proposed framework in the training and inferring process is comparable to those of deep learning-based road extraction algorithms.
5. The GEE and Google Colaboratory are proved to be ideal cloud-based platforms for large-scale remote sensing studies using deep learning algorithms.

The presented study well demonstrates the advantage of training deep neural networks on image samples with various land cover backgrounds adaptively for large-scale road extraction. The proposed framework can be also applied to other classification and object detection remote sensing tasks.

**Author Contributions:** Conceptualization, Y.L. and H.Z.; methodology, Y.L.; software, H.L.; validation, H.L. and Z.Y.; formal analysis, H.L.; investigation, H.L.; data curation, H.L.; writing—original draft preparation, Y.L. and H.L.; writing—review and editing, Z.Y., Y.Z. and H.Z.; visualization, H.L.; supervision, G.S.; project administration, G.S.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Key Research and Development Program of China (No.2021YFA075104), Scientific Research Project of Beijing Educational Committee (Grant no.: KM202110005024), the Natural Science Foundation of China (No. 41706201), and the Strategic Priority Program of the Chinese Academy of Sciences (XDB41020100).

**Data Availability Statement:** The OSM data is available at: <https://www.openstreetmap.org> (accessed on 15 August 2022); the Sentinel-2 data is provided by ESA at: <https://scihub.copernicus.eu/dhus/#/home> (accessed on 15 August 2022); the source code of the method proposed in this paper is available upon request.

**Acknowledgments:** The authors would like to thank the OpenStreetMap Project and ESA for providing the OSM street data and Sentinel-2 satellite images, respectively, and Google Inc. (Mountain View, CA, USA). for providing the Google Earth Engine and Google Colaboratory for processing of the large-scale remote sensing data. This work was supported in part by the National Key Research and Development Program of China (No. 2021YFA075104), Scientific Research Project of Beijing Educational Committee (Grant no.: KM202110005024), the Natural Science Foundation of China (No. 41706201), and the Strategic Priority Program of the Chinese Academy of Sciences (XDB41020100).

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

## References

1. Carrara, S.; Longden, T. Freight futures: The potential impact of road freight on climate policy. *Transp. Res. Part D Transp. Environ.* **2017**, *55*, 359–372. [\[CrossRef\]](#)
2. Xing, X.; Huang, Z.; Cheng, X.; Zhu, D.; Kang, C.; Zhang, F.; Liu, Y. Mapping human activity volumes through remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5652–5668. [\[CrossRef\]](#)
3. Tsou, J.Y.; Gao, Y.; Zhang, Y.; Sun, G.; Ren, J.; Li, Y. Evaluating urban land carrying capacity based on the ecological sensitivity analysis: A case study in Hangzhou, China. *Remote Sens.* **2017**, *9*, 529. [\[CrossRef\]](#)
4. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5891–5906. [\[CrossRef\]](#)
5. Zhang, B.; Wu, Y.; Zhao, B.; Chanussot, J.; Hong, D.; Yao, J.; Gao, L. Progress and challenges in intelligent remote sensing satellite systems. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1814–1822. [\[CrossRef\]](#)
6. Shi, S.; Zhong, Y.; Zhao, J.; Lv, P.; Liu, Y.; Zhang, L. Land-use/land-cover change detection based on class-prior object-oriented conditional random field framework for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 5600116. [\[CrossRef\]](#)
7. Buono, A.; Nunziata, F.; Mascolo, L.; Migliaccio, M. A multipolarization analysis of coastline extraction using X-band COSMO-SkyMed SAR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2811–2820. [\[CrossRef\]](#)
8. Bai, Y.; Sun, G.; Li, Y.; Ma, P.; Li, G.; Zhang, Y. Comprehensively analyzing optical and polarimetric SAR features for land-use/land-cover classification and urban vegetation extraction in highly-dense urban area. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102496. [\[CrossRef\]](#)
9. Zheng, K.; Wei, M.; Sun, G.; Anas, B.; Li, Y. Using vehicle synthesis generative adversarial networks to improve vehicle detection in remote sensing images. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 390. [\[CrossRef\]](#)
10. Li, H.; Li, S.; Song, S. Modulation Recognition Analysis Based on Neural Networks and Improved Model. In Proceedings of the 2021 13th International Conference on Advanced Infocomm Technology (ICAIT), Yanji, China, 15–18 October 2021. [\[CrossRef\]](#)
11. Tang, R.; Fong, S.; Wong, R.K.; Wong, K.K.L. Dynamic group optimization algorithm with embedded chaos. *IEEE Access* **2018**, *6*, 22728–22743. [\[CrossRef\]](#)
12. Bajcsy, R.; Tavakoli, M. Computer recognition of roads from satellite pictures. *IEEE Trans. Syst. Man Cybern.* **1976**, SMC-6, 623–637. [\[CrossRef\]](#)
13. Mena, J.B.; Malpica, J.A. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognit. Lett.* **2005**, *26*, 1201–1220. [\[CrossRef\]](#)
14. Kirthika, A.; Mookambiga, A. Automated road network extraction using artificial neural network. In Proceedings of the 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India, 3–5 June 2011. [\[CrossRef\]](#)
15. Simler, C. An improved road and building detector on VHR images. In Proceedings of the 2011 IEEE international geoscience and remote sensing symposium (IGARSS), Vancouver, BC, Canada, 24–29 July 2011. [\[CrossRef\]](#)
16. Shi, W.; Miao, Z.; Wang, Q.; Zhang, H. Spectral-spatial classification and shape features for urban road centerline extraction. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 788–792. [\[CrossRef\]](#)
17. Yin, D.; Du, S.; Wang, S.; Guo, Z. A direction-guided ant colony optimization method for extraction of urban road information from very-high-resolution images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4785–4794. [\[CrossRef\]](#)
18. 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. [\[CrossRef\]](#)
19. Lian, R.; Wang, W.; Mustafa, N.; Huang, L. 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. [\[CrossRef\]](#)
20. Mnih, V.; Hinton, G.E. Learning to detect roads in high-resolution aerial images. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010. [\[CrossRef\]](#)

21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015. [\[CrossRef\]](#)
22. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual U-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Zhou, L.; Zhang, C.; Wu, M. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018. [\[CrossRef\]](#)
25. Oehmcke, S.; Thrysøe, C.; Borgstad, A.; Salles, M.A.V.; Brandt, M.; Gieseke, F. Detecting hardly visible roads in low-resolution satellite time series data. In Proceedings of the 2019 IEEE International Conference on Big Data, Los Angeles, CA, USA, 9–12 December 2019. [\[CrossRef\]](#)
26. Zhou, G.; Chen, W.; Gui, Q.; Li, X.; Wang, L. Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5614115. [\[CrossRef\]](#)
27. Li, S.; Liao, C.; Ding, Y.; Hu, H.; Jia, Y.; Chen, M.; Xu, B.; Ge, X.; Liu, T.; Wu, D. Cascaded Residual Attention Enhanced Road Extraction from Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2021**, *11*, 9. [\[CrossRef\]](#)
28. Li, Y.; Zhu, M.; Sun, G.; Chen, J.; Zhu, X.; Yang, J. Weakly supervised training for eye fundus lesion segmentation in patients with diabetic retinopathy. *Math. Biosci. Eng.* **2022**, *19*, 5293–5311. [\[CrossRef\]](#)
29. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global land use/land cover with Sentinel 2 and deep learning. In Proceedings of the 2021 IEEE international geoscience and remote sensing symposium (IGARSS), Brussels, Belgium, 11–16 July 2021. [\[CrossRef\]](#)
30. Radoux, J.; Chomé, G.; Jacques, D.C.; Waldner, F.; Bellemans, N.; Matton, N.; Lamarche, C.; D’Andrimont, R.; Defourny, P. Sentinel-2’s Potential for Sub-Pixel Landscape Feature Detection. *Remote Sens.* **2016**, *8*, 488. [\[CrossRef\]](#)
31. Sun, G.; Liang, H.; Li, Y.; Zhang, H. Analysing the Influence of Land Cover Type on the Performance of Large-scale Road Extraction. In Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition, Shanghai, China, 15–17 October 2021. [\[CrossRef\]](#)
32. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [\[CrossRef\]](#)
33. Ebel, P.; Meraner, A.; Schmitt, M.; Zhu, X.X. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5866–5878. [\[CrossRef\]](#)
34. Venkatappa, M.; Sasaki, N.; Shrestha, R.P.; Tripathi, N.K.; Ma, H.-O. Determination of Vegetation Thresholds for Assessing Land Use and Land Use Changes in Cambodia using the Google Earth Engine Cloud-Computing Platform. *Remote Sens.* **2019**, *11*, 1514. [\[CrossRef\]](#)
35. Demetriou, D. Uncertainty of OpenStreetMap data for the road network in Cyprus. In Proceedings of the Fourth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2016), Paphos, Cyprus, 4–8 April 2016. [\[CrossRef\]](#)
36. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Groenen, P.J.F.; Kaymak, U.; van Rosmalen, J. Fuzzy clustering with minkowski distance functions. In *Advances in Fuzzy Clustering and Its Applications*; Valente de Oliveira, J., Pedrycz, W., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2007; pp. 53–68. [\[CrossRef\]](#)
38. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [\[CrossRef\]](#)
39. Krishna, K.; Murty, M.N. Genetic K-means algorithm. *IEEE Trans. Syst. Man Cybern. Part B* **1999**, *29*, 433–439. [\[CrossRef\]](#)
40. Ester, M. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, USA, 2–4 August 1996; pp. 226–231.
41. Ulku, I.; Akagündüz, E. A survey on deep learning-based architectures for semantic segmentation on 2D images. *Appl. Artif. Intell.* **2022**, *36*, 2032924. [\[CrossRef\]](#)
42. Nahler, G. Correlation Coefficient. In *Dictionary of Pharmaceutical Medicine*; Springer: Vienna, Austria, 2009; pp. 40–41. [\[CrossRef\]](#)