



Review Review of Road Segmentation for SAR Images

Zengguo Sun ^{1,2,*}, Hui Geng ², Zheng Lu ³, Rafał Scherer ⁴, and Marcin Woźniak ⁵

- ¹ Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, China
- ² School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; genghui@snnu.edu.cn
- ³ Institute of Remote Sensing Satellite, Beijing 100094, China; lvzheng@bit.edu.cn
- ⁴ Department of Intelligent Computer Systems, Czestochowa University of Technology, Armii Krajowej 36, 42-200 Częstochowa, Poland; rafal.scherer@pcz.pl
- ⁵ Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland; marcin.wozniak@polsl.pl
- * Correspondence: sunzg@snnu.edu.cn

Abstract: Road segmentation for synthetic aperture radar (SAR) images is of great practical significance. With the rapid development and wide application of SAR imaging technology, this problem has attracted much attention. At present, there are numerous road segmentation methods. This paper analyzes and summarizes the road segmentation methods for SAR images over the years. Firstly, the traditional road segmentation algorithms are classified according to the degree of automation and the principle. Advantages and disadvantages are introduced successively for each traditional method. Then, the popular segmentation methods based on deep learning in recent years are systematically introduced. Finally, novel deep segmentation neural networks based on the capsule paradigm and the self-attention mechanism are forecasted as future research for SAR images.

Keywords: synthetic aperture radar images; road segmentation; deep learning; capsule network; self-attention mechanism

1. Introduction

Synthetic aperture radar (SAR) is a kind of microwave sensor with an active working mode. The radar sensor emits energy pulse beams to the ground, and at the same time, receives a backscatter signal from the surface to detect the ground. As a new type of microwave imaging radar, SAR has many advantages [1,2]. It breaks through the limitation of optical remote sensing affected by weather and other external conditions. It has the ability to work all day and in all weather conditions and has rich characteristic signals, including amplitude, phase, polarization, and other information. Therefore, SAR is an essential remote sensing device for Earth observation. Since road segmentation for SAR images is closely related to important areas such as urban planning, residents' lives, and GIS database updates, it has always been a research hotspot in satellite image interpretation [3–5].

Because of the special imaging mode, SAR images have some features that optical images do not possess. The representation of the SAR image is not intuitive, and there are layover phenomena and speckle, which seriously affect the interpretation of images and then affect the extraction of road features [6]. Road detection is to draw a border around the road, predict its label, and only output whether there is a road and the road location information. However, road segmentation is to separate the road area from the background area according to the extracted image features. On the basis of the road detection output, it also predicts the road contour and further outputs the shape information, which makes road segmentation much more difficult than road detection.

In view of the above reasons, road segmentation for SAR images is a complex subject. In 1990, Samadani et al. [7] first proposed the method of local edge detection and then global road connection. Over the past 30 years, there has been in-depth research conducted on this topic and new methods and improvements are being constantly proposed. This paper



Citation: Sun, Z.; Geng, H.; Lu, Z.; Scherer, R.; Woźniak, M. Review of Road Segmentation for SAR Images. *Remote Sens.* 2021, *13*, 1011. https:// doi.org/10.3390/rs13051011

Academic Editor: Gilda Schirinzi

Received: 23 January 2021 Accepted: 4 March 2021 Published: 7 March 2021

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



Copyright: © 2021 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/). systematically summarizes road segmentation methods and related technologies for SAR images in recent years, points out the advantages and disadvantages of various methods, lists the comparison results between different algorithms, and finally proposes a new network architecture that combines the self-attention mechanism with the capsule network to form a new self-attention capsule network. The network can be combined with a neural network with a segmentation function to form a special road segmentation network suitable for SAR images.

The remainder of the paper is organized as follows. In Section 2, we review and discuss traditional road segmentation methods, divided into semi-automatic and automatic ones. Deep learning-based approaches are described in Section 3. Moreover, we compare some of them experimentally in Section 3.5. Finally, conclusions, discussion, and future prospects based on capsule networks and the attention mechanism are presented in Section 4.

2. Traditional Road Segmentation Methods of SAR Images

In order to facilitate the research on road segmentation algorithms for SAR images, the proposed algorithms are usually classified according to several aspects such as resolution size, processing flow, scene type, human intervention, and external auxiliary knowledge [8,9]. This paper mainly introduces the existing algorithms based on whether there is an artificial intervention or not.

2.1. Semi-Automatic Methods

The semi-automatic methods are based on human–computer interaction. Such methods often need to be provided with prior knowledge and then combine a road segmentation algorithm to achieve the goal, which involves many steps and high cost. Common methods include the active contour model (snake model), particle filter, template matching, mathematical morphology, extended Kalman filter (EKF), etc.

2.1.1. Snake Model

The basic principle of the snake model is to take control points that form a certain shape as the initial contour line. The contour line moves under the joint action of the internal force of the model and external force generated by image data and matches with the local features of an image to reach harmony so as to complete the segmentation of the image. In 2003, the snake model was used for SAR image road extraction for the first time [10]. The experimental results show that this model can be accurately fitted for straight or curved roads, but the initial contour line needs to be given for each road in the image, resulting in a large number of human–computer interactions. Fu Xiyou et al. [11] use a tensor voting algorithm that can extract significant structural features from an image masked by noise to obtain the curve saliency value for each point, and then uses the negative value of this value as the external energy of the snake model to extract roads from SAR images. However, the effect is not ideal for small roads. Saati et al. [12] combine multi-feature fusion with the snake model. This method increases the percentage and quality index of detected candidate roads, but it is sensitive to areas with backscatter similar to that of the roads.

2.1.2. Particle Filter

Particle filters are a nonlinear filtering method based on Monte Carlo. The basic idea of the Monte Carlo method is that when the problem to be solved is the probability of an event, or the expected value of a random variable, the frequency of the event or the average value of the random variable can be obtained by some experimental methods and used as the solution of the problem [13]. The concept of a particle filter is to express the distribution by random state particles extracted from the posterior probability. It is a sequential importance sampling method, generally divided into five steps: initialization, prediction, update, output, and resampling. Liu Junyi et al. [14] combine a particle filter with the snake model, select road seed points through a particle filter, and then use the snake model to connect

seed points to form a road. This method has high integrity and correctness even under the influence of high-backscatter objects near the road, but no topology relationship is considered when extracting complex road networks. Cheng Jianghua et al. [15] use the detected intersection as the starting point and track the centerline with a particle filter to extract roads, which overcomes the interference of various obstacles; additionally, parallel processing shortens the execution time of the extraction task. Mu Lin [16] adopts edge detection on the basis of filtering and mainly uses the Canny operator and a ratio of average (ROA) operator. This method improves the ROA edge detection operator to make its positioning more accurate. It also improves the Hough transform in the connection of line primitives, making it a straight-line detection technology in complex situations of large scenes.

2.1.3. Template Matching

Template matching is to find a region that matches the given sub-image in the entire image region. The working principle is that given a template image and an image to be detected, the matching degree between the template image and the overlapping subimage is calculated from left to right and from top to bottom in the image to be detected. The greater the matching degree, the greater the likelihood that the two will be the same. Cheng Jianghua et al. [17] propose a method based on a circular template matching. Firstly, two points are input to calculate a circular template and road direction, then the template is matched with the image on the considered road direction to look for center points, and finally, the extracted center points are connected by the conic fitting. Su Yang et al. [18] start from the model of single lanes and isolation belts in highways, use circular template matching method to extract the centerline, and finally extract the entire highway based on the width of single lanes. This method can eliminate the influence of noise and is able to extract the highway completely. Han Ping et al. [19] propose a multi-stage classification algorithm for runway detection in polarimetric SAR images. The prior information, statistical characteristics of the polarization coherence matrix, and a total polarization power detector are used to complete the three-level classification, and then runway areas are extracted.

2.1.4. Mathematical Morphology

The basic operations of mathematical morphology mainly include erosion and dilation and open and closed operation, which are applicable to all aspects related to image processing. Yu Jie et al. [20] propose a new road network extraction method based on statistical characteristics and road shape features of SAR images. This method solves the problem of road width changes in high-resolution images and reduces the influence of detailed information but cannot reduce the influence of strong scatterers in the road. Xiao Hongguang et al. [21] use parametric kernel graph cuts to perform primary segmentation of road targets in high-resolution SAR images, fill holes with mathematical morphology to extract the centerline of road targets, and restore road width to obtain satisfactory road extraction results. This method omits preprocessing and reduces time cost. Lu Xiaoguang et al. [22] propose an adaptive unsupervised classification method for runway detection in polarimetric synthetic aperture radar (PolSAR) images. This method can quickly and accurately detect runways and has good robustness. Filippo Biondi [23] proposes an improved full-polarization SAR decomposition scheme, which uses Doppler sub-aperture multi-chromatic analysis to achieve more accurate classification. This method produces significantly improved results.

Generally, these methods are used in nearly all computer vision and image classification methods to improve the results.

2.1.5. Extended Kalman Filtering

The Kalman filter obtains the optimal estimation of the system state through recursive processing of probability density functions, but it can only be processed on linear systems. Extended Kalman filtering (EKF) generalizes the Kalman filter to nonlinear systems through

local linearization, which realizes the extraction of nonlinear targets. Zhao Jinqi et al. [24] propose an algorithm based on EKF and the particle filter by analyzing road characteristics. The algorithm is suitable for medium- and high-noise road scenes, but it cannot automatically switch the thresholds of Kalman and particle filters according to the actual situation. Yu Jie et al. [25] combine the improved profile matching algorithm with EKF to effectively extract roads in complex scenes with less manual intervention, but the accuracy of road extraction is not high in corners and areas with weak scattering features.

In the above semi-automatic methods, the snake model can fit straight and curved roads better, EKF is suitable for nonlinear systems, and the particle filter extends the extraction range to non-Gaussian systems in the former. Template matching has a certain effect in eliminating the influence of noise and interference, and mathematical morphology can mainly simplify image data. However, these methods need data inputted manually, and excessive human–computer interaction reduces the algorithm's efficiency.

2.2. Automatic Methods

Automatic methods do not require manual intervention and mainly depend on the selection of road features in positioning. Due to the complexity of road scenes and a large amount of surface interference information, current algorithms are only suitable for the automatic extraction of a certain type or a specific scene, which cannot meet the requirements for automatic extraction of all scenes. Common methods include dynamic programming, Markov random field (MRF) models, genetic algorithms (GAs), and fuzzy connectedness.

2.2.1. Dynamic Programming

The principle of dynamic programming is to set two road edges as the starting points based on the radiation or geometric characteristics of road line primitives and search for the next potential edge point of the road within a certain sector according to the principle of the minimum cost function. Jia Chengli et al. [26] use a detection operator to detect road edges, then employ a series of templates to calibrate edge pixels and connect short line segments, and finally use dynamic programming techniques to connect road curve segments. This method can extract most of the roads in the image, but there are still fractures and false alarms. Hong Richang et al. [27] propose a method based on edge line segmentation grouping and dynamic programming to track line segments, which has a good recognition effect on complex urban road networks and mountain roads with large background interference in SAR images but cannot effectively achieve multi-scale automatic road recognition. He Chu et al. [28] propose a method based on compressed sensing and a multi-scale pyramid. This method can not only take advantage of the observation matrix to reduce the feature space dimensionality but also analyzes the texture features of the polarization interferometric SAR image at different scales.

2.2.2. Markov Random Field Model

The Markov random field (MRF) model can make full use of context information and prior knowledge of image features and often shows a better connection effect. However, the road extraction algorithms based on the MRF model usually have the disadvantages of slow iteration speed and inability to meet real-time requirements. Tupin et al. [29] apply the MRF model to the problem of SAR image road extraction for the first time. The idea is to build an MRF model according to the length and angle rules and abstract the problem of road network global connection into the problem of solving the maximum posterior probability of total potential energy. Chen Lifu et al. [30] propose an algorithm combining MRF segmentation and mathematical morphology processing, which uses an MRF segmentation algorithm based on an iterative conditional mode (ICM) algorithm to segment SAR images. Then they use multiple factors to remove false alarms to obtain road targets according to the geometric characteristics of the road. Cheng Jianghua et al. [31] state that the traditional MRF-based algorithms usually require a large number of calculation

operations, which are relatively time-consuming and difficult to apply. Therefore, a GPUaccelerated road extraction method based on MRF is proposed, which effectively improves the calculation efficiency.

2.2.3. Genetic Algorithms

The main feature of genetic algorithms is to directly operate on the structure itself, which has the advantages of high parallelism and strong global search ability. However, there are too many parameters to be set, and the parameter selection depends on experience values, so the practicability is not wide. Jiang Yunhui et al. [32] first filter the SAR image twice to obtain a thin road centerline, then perform the Hough transform on the binary image to obtain road segments, and finally use genetic algorithms to connect the roads. In [33], a genetic algorithm is used to connect line primitives after line feature detection and line primitive extraction, and the effect of this method is promising. The main road extraction algorithm proposed by Xiao Qiangzhi et al. [34] is to cluster the filtered image first, then build a road model, and finally use a genetic algorithm to search for the global optimal road. This method has the advantages of fewer manual setting parameters and a faster calculation speed, but it is not suitable for the extraction of complex roads.

2.2.4. Fuzzy Connectedness

Fuzzy connectedness theory uses "fuzzy similarity" to describe the similarity between pixels. The targets it recognizes are consistent with the characteristics of road network objects, so it is suitable for automatic recognition of the road network information. Udupa et al. [35] first propose an image segmentation method that uses fuzzy connectedness to describe the tightness of different pixels, and it has been widely used. The traditional fuzzy connectedness theory needs to define the starting point of the road object clearly in the image, which greatly reduces the efficiency and feasibility. Fu Xiyou et al. [36] automatically obtain seed points with high confidence by using the detection results of the ratio of exponentially weighted averages (ROEWA) operator and fuzzy c-means road segmentation results. Then they use the fuzzy connectedness algorithm to expand seed points to extract roads and obtain final results after morphological processing. This method can effectively extract roads with different widths and bends without manual input of seed points.

In the above automatic methods, a dynamic programming algorithm uses radiation or geometric features of line elements, the MRF model mainly uses context information and prior knowledge of image features, genetic algorithms are a global search method, and the fuzzy connectedness algorithm is a region-based segmentation method. Although these algorithms have reduced the number of human–computer interactions, they are still not fully automated.

2.3. Advantages and Disadvantages of Traditional Road Segmentation Methods

There are many kinds of traditional road segmentation algorithms for SAR images, most of which have better performance in solving specific categories and have improved greatly in efficiency. However, the processing procedures of these methods are relatively complicated and involve many steps. When the same algorithm is used in different scenes, there are great differences in the extraction effect and accuracy. By analyzing the characteristics of the above algorithms, it can be seen that traditional methods belong to the model-driven methods, i.e., they rely heavily on a specific model and then on specific assumptions. Therefore, the adaptability and stability of such methods are generally not strong.

In summary, the advantages and disadvantages of several typical traditional segmentation algorithms are shown in Table 1.

Methods	Advantages	Disadvantages
Snake model	Good fitting effect on straight lines and curve contours	Setting initial contour curves for each road
Particle filter	Suitable for all nonlinear and non-Gaussian systems	Dependent on the estimation of initial state
Template matching	Fewer effects of noise and interference	Dependent on the actual situation to choose threshold
Mathematical morphology	Simplified image data	Easily appearing fracture
Extended Kalman filtering (EKF)	Suitable for nonlinear systems	Large error and filter divergence
Dynamic programming	Simple algorithm and low computational complexity	Not suitable for the case of large spacing between primitives
Markov random field (MRF)	Makes full use of image context information and prior knowledge	No real-time and slow iterative connection
Genetic algorithms (GA)	Strong adaptability and good connection effect	Many parameters and dependent on experience values
Fuzzy connectedness	Good description of fuzzy areas in the image Complex computation and weak effectiveness	

Table 1. Advantages and disadvantages of traditional road segmentation methods.

3. Road Segmentation Methods for SAR Images Based on Deep Learning

3.1. Background of Deep Learning Methods

With the rapid development of SAR satellites and imaging technology, more and more satellites can provide continuous and more reliable ground observation data. Moreover, their return period is constantly shortened, and they can continuously observe for a long time. Therefore, they can provide massive data, which shows that the SAR big data era has now begun. With the rapid growth of the data scale, traditional model-driven methods are gradually becoming unable to meet the needs of big data applications. So, intelligent processing methods represented by deep learning have emerged. Such methods show excellent results in natural image processing and good capabilities in remote sensing.

3.2. The Development of Target Detection Networks

Deep learning aims to perform automatic extraction of multi-layer feature representations from data [37–39], which has been successfully applied to target recognition. In 2014, Girshick et al. proposed the Region Convolutional Neural Network (R-CNN) model [40], which uses the selective search algorithm [41] to extract regional candidate boxes. Although the performance of this algorithm has been greatly improved, there are also many problems, such as complicated steps and a large number of calculations, which restrict the performance of the algorithm. In response to the shortcomings of R-CNN, He et al. [42] equip the networks with spatial pyramid pooling and name the new network structure SPP-Net. This network performs only one convolution operation, which greatly reduces the amount of calculation, but the network is divided into multiple stages during training and still depends on the generation of candidate regions. Fast R-CNN [43] borrows the idea of SPP-Net on the basis of R-CNN, which improves the detection accuracy and speed at the same time. However, the network still consumes much time in extracting candidate regions, which cannot meet the real-time requirements of the algorithm. Ren et al. [44] propose Faster R-CNN to solve the above problem of the slow network running speed. The network uses a region proposal network (RPN) instead of the selective search algorithm and is superior to the single-stage detection network in controlling the proportion of positive and negative samples and adjusting the candidate frame position more precisely [45]. However, RPN uses anchor points of different scales, which may cause the problem of variable target size and inconsistent receptive fields when mapping to the original image. The Regionbased Fully Convolutional Network (R-FCN) [46] follows the Faster R-CNN framework

and uses a fully convolutional neural network, but the algorithm still involves a large amount of computation and it is difficult to meet the real-time requirement. The above networks are all used in the detection problem and cannot segment objects such as roads.

In 2016, Multi-task Network Cascades (MNC) [47] was proposed, which divides the semantic segmentation task into three parts, namely, differentiating instances, estimating masks, and categorizing objects. On the basis of shared features, the three tasks are performed simultaneously and independently of each other, and the output of the previous task is used as the input of the next task, thus forming a hierarchical multi-task structure. Fully Convolutional Instance-aware Semantic Segmentation (FCIS) [48] is the first fully convolutional end-to-end solution for instance-aware semantic segmentation tasks. It can detect and segment multiple instances at the same time, and introduce position-sensitive inside/outside score maps to realize the fully sharing of the underlying convolution representation between the two sub-tasks, as well as between all regions of interest. In 2017, Mask R-CNN was proposed, which outperformed all existing, single-model solutions at that time in all tasks, including MNC and FCIS. A target mask module is added to this network on the basis of Faster R-CNN, and the network framework is shown in Figure 1. Mask R-CNN is mainly composed of three parts of the network: the RPN network part using the convolutional neural network to extract the feature map, the network part to generate the target classification using region proposals, and the network part for semantic segmentation and mask generation [49]. The steps of the algorithm are as follows. Firstly, images are input into a deep convolutional network to obtain the feature map, and then a set of rectangular target frames and their corresponding target scores are obtained using RPN.



Figure 1. Mask Region Convolutional Neural Network (R-CNN) network structure.

After that, the region of interest (ROI) is further processed using the ROI alignment method. Finally, these converted proposed regions are passed to the classifier to output the bounding boxes of the corresponding roads, and the semantic segmentation network part generates road masks in parallel. Mask R-CNN is suitable for pixel segmentation. It can not only classify and locate the target box but can also perform fine-grained segmentation of objects such as roads, which has excellent flexibility.

In summary, the performance of each network and whether it can conduct instance segmentation are shown in Table 2.

Networks	Year Proposed	Features	Shortcomings	Instance Segmentation
R-CNN	2014	Adding region proposal	Cumbersome steps and slow speed	No
SPP-Net	2015	Convolution once and adding spatial pyramid pooling	Dependent on generation of candidate regions	No
Fast R-CNN	2015	Adding ROIPooling and using multi-task loss function	No real time	No
Faster R-CNN	2015	Replacing selective search with RPN	Variable target size and inconsistent feeling field	No
R-FCN	2016	Improving ROIPooling and adopting ResNet in backbone network	Much calculation and no real time	No
MNC	2015	Hierarchical multi-tasking structure, and sharing underlying convolution features	Great loss of detail information, and over-parameterization	Yes
FCIS	2016	Adding position-sensitive inside/outside score maps, and end-to-end solution	Appearance of systematic artifacts on overlapped objects	Yes
Mask R-CNN	2017	Adding ROIAlign and semantic segmentation branch	Low feature representation ability, and missing spatial feature information	Yes

Table 2. Comparison of neural network-based methods.

3.3. Deep Learning Methods

In recent years, there have been many research results that use deep learning methods to solve road segmentation problems for SAR images. In 2018, Henry et al. [50] used a fully convolutional neural network (FCNN) model to segment roads in TerraSAR images. This method can separate thin objects and detect a variety of road patterns in speckle environments, but it has a poor prediction effect on interference objects of the forest boundary type. At the same time, the learning features of the traditional FCNN are usually high-dimensional and take up a lot of computing resources. In 2019, Chen Hua et al. [51] proposed a new recognition method for solving the problems in [50]. This method improves the FCNN and moves convolutional layers backward to enhance the expression of the final inverse convolutional layer and reduce information loss. The method is effective for SAR images and has achieved good results. Due to the special coherent imaging mode, speckle appears in SAR images, which makes the interpretation of SAR images very difficult. In addition, in the segmentation technology, sample labeling accuracy is very high and, currently, free and public road segmentation datasets for SAR images are scarce. These factors have seriously affected the development of deep learning methods in road segmentation for SAR images, so there are few related works on deep learning. Compared with SAR images, general optical remote sensing images are easier to process, and there are many deep learning algorithms for road segmentation of such images. In [52], a deconvolution neural network is used to initially segment the road scene, and then the final result is obtained by further processing based on color and depth information. This method has a good segmentation effect at the boundary between classes but still needs to use a larger dataset for evaluation. Li Haoyu et al. [53] propose a deep learning road extraction model based on a similarity mapping relationship. This model directly stores knowledge in the network instead of just learning a set of feature extraction and integrated network parameters. Cheng Guangliang et al. [54] cascade a road detection network and a centerline extraction network into a framework and train the proposed new network through an end-to-end strategy. This method is able to obtain a smooth and complete centerline, but it cannot deal with shaded regions well.

As we all know, training plays a critical role in deep learning methods. Moreover, there are many aspects that need to be considered in training, such as training examples,

loss functions, and convergence. For training examples, their sources should be wide, the types should be rich, and they should contain as many situations as possible. This guarantees better robustness of the model. In addition, the samples are usually cropped to a uniform size, and the cropped examples cannot be distorted. For loss function, common loss functions used for segmentation networks include cross-entropy loss, focal loss, dice loss, intersection over union (IOU) loss, Tversky loss, and so on. Among them, the cross-entropy loss can be used in most segmentation scenes, but the effect is not good when the number of current scene pixels is less than the number of background pixels. Focal loss is proposed to solve the problem of the imbalance in the number of difficult and easy samples and is mainly used in the two-classification situation. Dice loss is used when the number of positive and negative samples is extremely unbalanced, and it may affect backpropagation if used under normal circumstances. For convergence, the convergence can be reflected by the change of the loss function value. The loss function calculates the error between the forward calculation result of each iteration of the neural network and the true value. Then, according to the derivative of the loss function, the error is propagated back along the minimum gradient direction to update each weight value in the forward calculation process and, finally, the iteration is stopped when the loss function value tends to a satisfactory value. At this moment, convergence is achieved, and the optimal weight coefficients are obtained.

3.4. Advantages and Disadvantages of Deep Learning Methods

Due to the emergence and wide application of big data, deep learning methods have emerged and achieved an excellent development level. They are typical data-driven methods, and any desired results can be obtained theoretically when there are enough training data. Unlike the traditional methods, deep learning methods no longer depend on specific models and constraints and can construct feature extractors adaptively according to the training data. In addition, the feature extractor and the classifier can carry out endto-end training as a whole, avoiding the complex steps of data modeling, feature design, and classifier selection in traditional methods, making the processing flow more convenient and efficient. However, the methods are too dependent on data, and thus require huge sample sets for training, but the labeling of road samples with specific shapes is timeconsuming and laborious work. This is a big disadvantage of deep learning methods. Furthermore, because of the need to train the network with massive data, the hardware requirements are also very high. Moreover, deep learning methods also have disadvantages, such as the inability to judge the correctness of the data and to modify the learning results easily, a large amount of calculation, and low interpretability and explainability.

3.5. Performance Comparison of Common Algorithms

In order to verify the effect of deep learning methods on road segmentation, three common segmentation algorithms, Mask R-CNN, FCIS, and MNC, with their standard settings, are trained on the road dataset. This dataset contains 10,026 image chunks from 23 scenes of GF-3 SAR images, and each chunk has a pixel size of 512×512 . The imaging modes include Spotlight (SL), Ultra-Fine Strip (UFS), Fine Strip I (FSI), and Fine Strip II (FSI), and the corresponding resolution is 1 m, 3 m, 5 m, and 10 m, respectively. The operating system of the experimental machine is Ubuntu 16.04, and the GPU is NVIDIA 2080ti. The results are shown in Table 3. Average precision (*AP*) and intersection over union (*IoUI*) are used to measure the segmentation performance of each algorithm and their calculation formulas are as follows.

$$AP = \sum_{0}^{1} (r_{n+1} - r_n) \max_{\tilde{r}:\tilde{r} \ge r_{n+1}} p(\tilde{r})$$
(1)

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$
(2)

where $p(\tilde{r})$ is the measured precision at recall \tilde{r} , B_P is the predicted road mask, B_{gt} is the road label frame, $area(B_p \cap B_{gt})$ is the area of the intersection of the predicted and ground truth bounding boxes, and $area(B_p \cup B_{gt})$ is the area of their union. Here, the precision p is often used to reflect the correct rate of a category being correctly predicted, and the recall r is used to reflect the proportion of correctly predicted samples among all predicted samples. The higher the value of AP and IoU, the better the algorithm performance.

Table 3. Comparison of road segmentation performance among deep learning algorithms.

Algorithms	Average Precision (AP) (%)	Intersection over Union (IoU) (%)
Mask R-CNN	86.5	88.2
FCIS	79.2	85.7
MNC	73.1	80.0

As can be seen from Table 3, Mask R-CNN has the best performance, and its *AP* and *IoU* are significantly better than FCIS and MNC. By analyzing the structures of three systems, it can be found that Mask R-CNN replaces ROIPooling with ROIAlign to extract more accurate features, which is one of the reasons why its performance is better than other algorithms.

In order to further verify the performance of Mask R-CNN, Figures 2–4 show its segmentation results on three road shapes: straight, three-fork, and "V". Each time, 500 samples are randomly selected from the training set and added to the training data. The final network model after the last training is used as the initial network model for the next training. It can be seen from the figures that with the increase in training samples, Mask R-CNN segmentation is more accurate.



Figure 2. Mask R-CNN algorithm segmentation results of straight road: (**a**) test image, (**b**) 500 training samples, (**c**) 1000 training samples, (**d**) 1500 training samples, (**e**) 2000 training samples.



Figure 3. Mask R-CNN algorithm segmentation results of three-fork road: (**a**) test image, (**b**) 500 training samples, (**c**) 1000 training samples, (**d**) 1500 training samples, (**e**) 2000 training samples.



Figure 4. Mask R-CNN algorithm segmentation results of V-shaped road: (**a**) test image, (**b**) 500 training samples, (**c**) 1000 training samples, (**d**) 1500 training samples, (**e**) 2000 training samples.

4. Conclusions and Future Prospects

4.1. Conclusions

Road segmentation for SAR images plays an important role in the field of remote sensing. At present, the research on this topic has made great progress. However, due to the complex background of SAR images and the influence of speckle, it is still difficult to extract road features in SAR images. This paper systematically summarizes the research achievements in this field in recent years. According to algorithms' characteristics, they are divided into traditional road segmentation methods and road segmentation methods based on deep learning. The traditional segmentation methods are further divided into two categories: semi-automatic and automatic, according to the degree of automation. The traditional road segmentation methods are model-driven methods, which need to build models and design features according to the prior knowledge, and then determine parameters of the corresponding model. It should be noted that traditional methods have several problems, such as over-dependence on model parameters, too complex model structure, and low prediction accuracy. In addition, an unreasonable feature design usually results in a weak feature representation. The methods based on deep learning are data driven. This kind of method no longer relies on specific models or assumptions but starts from image data itself to find an internal connection mechanism between them and uses a large amount of data to train the automatic learning features. However, most of the existing networks cannot deal with high-dimensional complex images like SAR images well and require a large number of labeled datasets for training. Therefore, the accuracy and efficiency of deep learning methods still have huge room for improvement.

4.2. Future Prospects

The road segmentation methods based on deep learning have greatly improved in performance. Most of the networks mentioned are developed on the basis of CNNs. Yet, CNNs have the following drawbacks. Firstly, a CNN's scalar output makes the network's feature representation capability low. Secondly, a CNN uses a pooling operation that may discard information about the precise location of entities in the region. These factors will affect the accuracy of road segmentation networks. In order to overcome such shortcomings of CNNs, CapsNet is proposed [55]. It uses the dynamic routing mechanism to replace the pooling operation, which can extract spatial feature information of data well and identify objects that are not easy to detect in training data from different perspectives. This network shows its unique advantages in remote sensing image detection [56,57], general optical image detection [58], emotional analysis [59], speech analysis [60], and text analysis [61]. However, CapsNet does not involve local constraints of feature learning and is not suitable for selecting local features. For images with complex backgrounds, the performance of CapsNet needs to be improved. On the other hand, the computational cost of the dynamic routing process is very high, and it will produce higher memory requirements as the feature dimension increases. The self-attention mechanism [62] calculates the response of a certain position as the weighted sum of features of all positions, reduces the dependence on external information, and is better at capturing the internal correlation of data or features, which can help the model focus on more relevant regions in the image. As a nonlocal operation, it solves the task of learning important features when CapsNet considers positions and obtains better classification performance in the case of fewer data samples or a more complex image background [63–65]. The introduction of a self-attention mechanism can solve the special problems encountered by the network when processing information locally, and the output of each activation is modulated by a subset of other activations, which helps the network consider smaller parts of the image when necessary. Moreover, the self-attention mechanism provides better classification with a lower computational cost. Therefore, the self-attention mechanism and CapsNet can be combined to form a self-attention capsule network. Firstly, the information extracted from the initial convolution layer of the capsule network is input into the self-attention mechanism to generate the self-attention graph, which helps to eliminate the ambiguity of uncorrelated and noise response. Secondly, the main features are input into the main capsule layer, and then input into the classification layer. The improved network uses a relatively shallow CapsNet architecture to reduce the computational load and uses a self-attention module to compensate for the lack of a deep network, thereby significantly improving CapsNet's local feature selection ability. The combination of self-attention capsule network and neural network with segmentation function can form a special road segmentation network suitable for SAR images. This network model can be processed in parallel and reduce the training time during network training and the amount of calculation. Moreover, under the premise of fully extracting the spatial relationship of the data, the salient features useful for specific tasks can be highlighted so as to improve the adaptability and accuracy of the road segmentation network. This can be an important direction for future research. However, the model design is slightly complicated and may not be easy to understand. In addition, the performance of the self-attention capsule network combined with different segmentation networks will be very different, which requires repeated comparisons through a large number of experiments. At present, it is just a network structure design, and the following work can focus on network optimization and other aspects.

Author Contributions: Conceptualization and methodology, Z.S.; formal analysis and investigation, H.G.; data curation, Z.L.; writing—original draft preparation, H.G.; writing—review and editing, R.S. and M.W.; supervision, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61102163), the Fundamental Research Funds for the Central Universities (No. GK201903085), the Key Laboratory of Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of the People's Republic of China (No. KLSMNR-202004) and the State Key Laboratory of Geo-Information Engineering (No. SKLGIE2019-M-3-5).

Acknowledgments: The authors would like to thank the China Center for Resources Satellite Data and Application for providing SAR images.

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

References

- 1. Zhang, Y.N.; Li, Y. The Key Technology of SAR Image Processing; Electronic Industry Press: Beijing, China, 2004.
- 2. Oliver, C.; Quegan, S. Understanding Synthetic Aperture Radar Images; Artech House: Boston, MA, USA, 2004.
- Cui, B.; Zhang, Y.H.; Yan, L. A SAR change detection method based on the consistency of single-pixel difference and neighborhood difference. *Remote Sens. Lett.* 2019, 10, 488–495. [CrossRef]
- 4. Han, Z.-S.; Wang, C.-P.; Fu, Q. Arbitrary-oriented target detection in large scene sar images. *Def. Technol.* **2020**, *16*, 933–946. [CrossRef]
- Yang, F.; Wang, H.; Jin, Z. A fusion network for road detection via spatial propagation and spatial transformation. *Pattern Recognit.* 2020, 100, 107141. [CrossRef]
- Yu, Z.; Wang, W.; Li, C.; Liu, W.; Yang, J. Speckle Noise Suppression in SAR Images Using a Three-Step Algorithm. Sensors 2018, 18, 3643. [CrossRef]
- Samadani, R.; Vesecky, J. Finding Curvilinear Features in Speckled Images. IEEE Trans. Geosci. Remote. Sens. 1990, 28, 669–673. [CrossRef]
- 8. Cheng, J.H.; Gao, G.; Ku, X.S. Review of road network extraction from SAR images. J. Image Graph. 2013, 18, 11–23.

- Cheng, J.H. Road Extraction in High-Resolution SAR Images; National University of Defense Technology: Changsha, China, 2013; pp. 1–132.
- 10. Bentabet, L.; Jodouin, S.; Ziou, D.; Vaillancourt, J. Road vectors update using SAR imagery: A snake-based method. *IEEE Trans. Geosci. Remote Sens.* 2003, *41*, 1785–1803. [CrossRef]
- 11. Fu, X.Y.; Zhang, F.L.; Wang, G.J. SAR image road extraction combining tensor voting and Snakes model. *J. Image Graph.* **2015**, *20*, 1403–1411.
- 12. Saati, M.; Amini, J. Road Network Extraction from High-Resolution SAR Imagery Based on the Network Snake Model. *Photogramm. Eng. Remote. Sens.* 2017, *83*, 207–215. [CrossRef]
- 13. Nicholas, M.; Stanislaw, U. The Monte Carlo Method. J. Am. Stat. Assoc. 1949, 44, 335–341.
- 14. Liu, J.; Sui, H.; Tao, M.; Sun, K.; Mei, X. Road extraction from SAR imagery based on an improved particle filtering and snake model. *Int. J. Remote. Sens.* **2013**, *34*, 8199–8214. [CrossRef]
- 15. Cheng, J.H.; Gao, G. Parallel particle filter for tracking road centerlines from high-resolution SAR images using detected road junctions as initial seed points. *Int. J. Remote Sens.* **2016**, *37*, 4979–5000. [CrossRef]
- Mu, L. Study on Road Extraction of High Resolution Polarimetric SAR Interferometry Data; Liaoning University of Engineering and Technology: Fuxin, China, 2010; pp. 1–90.
- Cheng, J.H.; Guan, Y.F.; Ku, X.S. Semi-automatic road centerline extraction in high-resolution SAR images based on circular template matching. In Proceedings of the International Conference on Electric Information and Control Engineering, Wuhan, China, 15–17 April 2011; pp. 1688–1691.
- 18. Su, Y.; Shen, T. High-resolution SAR image highway extraction algorithm. Remote Sens. Inf. 2012, 27, 8–13.
- 19. Han, P.; Xu, J.S.; Zhao, A.J. PolSAR image runways detection based on multi-stage classification. *J. Syst. Eng. Electron.* **2014**, *36*, 866–871.
- 20. Yu, J.; Liu, Z.Y.; Yan, Q. High-resolution SAR image road network extraction combining statistics and shape features. *J. Wuhan Univ.* **2013**, *38*, 1308–1312.
- 21. Xiao, H.G.; Wen, J.; Chen, L.F. New road extraction algorithm of high-resolution SAR image. Comput. Eng. Appl. 2016, 52, 198–202.
- 22. Lu, X.G.; Lin, Z.S.; Han, P. Fast detection of airport runways in PolSAR images based on adaptive unsupervised classification. *J. Remote Sens.* **2019**, 23, 1186–1193.
- 23. Biondi, F. Multi-chromatic analysis polarimetric interferometric synthetic aperture radar (MCA-PollnSAR) for urban classification. *Int. J. Remote Sens.* 2019, 40, 3721–3750. [CrossRef]
- 24. Zhao, J.Q.; Yang, J.; Li, P.X. Semi-automatic road extraction from SAR images using EKF and PF, ISPRS-International Archives of the Photogrammetry. *Remote Sens. Spat. Inf. Sci.* 2015, *7*, 227–230.
- 25. Zhao, J.Q.; Yang, J.; Li, P.X. Semi-automatic road extraction from SAR images using an improved profile matching and EKF. *Geomat. Inf. Science Wuhan Univ.* **2017**, *42*, 1144–1150.
- 26. Jia, C.L.; Kuang, G.Y. Automatic extraction of roads from low resolution SAR images. J. Image Graph. 2005, 10, 1218–1223.
- Hong, R.C.; Wu, X.Q.; Liu, Y. Research on roads automatic extraction from low resolution remote sensing image. *J. Remote Sens.* 2008, 12, 36–45.
- He, C.; Liu, M.; Feng, Q. PolInSAR Image Classification Based on Compressed Sensing and Multi-scale Pyramid. *Acta Autom. Sin.* 2011, 37, 820–827.
- 29. Tupin, F.; Maitre, H.; Mangin, J.-F.; Nicolas, J.-M.; Pechersky, E. Detection of linear features in SAR images: Application to road network extraction. *IEEE Trans. Geosci. Remote. Sens.* **1998**, *36*, 434–453. [CrossRef]
- 30. Chen, L.F.; Wen, J.; Xiao, H.G. A road extraction algorithm combining MRF segmentation and mathematical morphology. *China Space Sci. Technol.* **2015**, *35*, 17–24.
- Cheng, J.H.; Ding, W.X.; Zhu, X.W. GPU-accelerated main road extraction in Polarimetric SAR images based on MRF. In Proceedings of the 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 928–932.
- 32. Jiang, Y.H.; Pi, Y.J. SAR image road detection based on Hough transform and genetic algorithm. *Radar Sci. Technol.* **2005**, *3*, 156–162.
- 33. Jia, C.J.; Zhao, L.J.; Wu, Q.C. Automatic road extraction from SAR imagery based on genetic algorithm. *J. Image Graph.* **2008**, *13*, 1134–1142.
- 34. Xiao, Q.Z.; Bao, G.S.; Jiang, X.Q. Road network extraction in classified SAR images using genetic algorithm. *J. Image Graph.* 2004, 93, 93–98. [CrossRef]
- 35. Udupa, J.K.; Samarasekera, S. Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation. *Graph. Model. Image Process.* **1996**, *58*, 246–261. [CrossRef]
- 36. Fu, X.Y.; Zhang, F.L.; Wang, G.J. Automatic road extraction from high resolution SAR images based on fuzzy connectedness. *J. Comput. Appl.* **2015**, *35*, 523–527.
- 37. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117. [CrossRef]
- 38. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* 2015, 521, 436–444. [CrossRef]
- 39. Zhou, F.Y.; Jin, L.P.; Dong, J. Review of research on convolutional neural networks. *Chin. J. Comput.* 2017, 40, 1229–1251.
- 40. Girshick, R.; Donahue, J.; Darrell, T. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
- 41. Uijlings, J.R.R.; Sande, K.E.A.; Gevers, T. Selective search for object recognition. Int. J. Comput. Vis. 2013, 104, 154–171. [CrossRef]

- 42. He, K.M.; Zhang, X.Y.; Ren, S.P. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
- Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- 44. Ren, S.P.; He, K.M.; Girshick, R. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
- 45. Sun, X.; Wang, Z.R.; Sun, Y.R. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. J. Radars 2019, 8, 852–862.
- Dai, J.F.; Li, Y.; He, K.M.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. *Neural Inf. Process. Syst.* 2016, 179–387. [CrossRef]
- Dai, J.F.; He, K.M.; Sun, J. Instance-aware Semantic Segmentation via Multi-task Network Cascades. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1–10.
- Li, Y.; Qi, H.Z.; Dai, J.F. Fully Convolutional Instance-aware Semantic Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 1–9.
- He, K.M.; Gkioxari, G.; Doll, R.P. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
- 50. Corentin, H.; Majid, A.S.; Nina, M. Road segmentation in SAR satellite images with deep fully convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1867–1871.
- 51. Chen, H.; Guo, W.; Yan, J.W. New method of SAR image road recognition based on deep learning. J. Jilin Univ. 2020, 50, 1778–1787.
- 52. John, V.; Guo, C.Z.; Mita, S. Fast road scene segmentation using deep learning and scene-based models. In Proceedings of the International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 3763–3768.
- Li, H.Y.; Chen, Y.P.; Yang, Y. Deep learning road extraction model based on similarity mapping relationship. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9799–9802.
- 54. Cheng, G.L.; Wang, Y.; Xu, S.B. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 3322–3337. [CrossRef]
- 55. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1–11.
- Islam, K.A.; Pérez, D.; Hill, V. Seagrass detection in coastal water through deep capsule networks. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, 23–26 November 2018; Springer: Cham, Switzerland, 2018; pp. 320–331.
- 57. Singh, M.; Singh, R.; Vatsa, M. Dual directed capsule network for very low-resolution image recognition. In Proceedings of the International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 1–10.
- Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P. Synthetic aperture radar ship detection using capsule networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 725–728.
- Wang, Y.Q.; Sun, A.X.; Han, J.L. Sentiment analysis by capsules. In Proceedings of the Web Conference, Lyons, France, 23–27 April 2018; pp. 1165–1174.
- 60. Iqbal, T.; Xu, Y.; Kong, Q.Q. Capsule routing for sound event detection. In Proceedings of the European Signal Processing Conference, Rome, Italy, 3–7 September 2018; pp. 1–5.
- 61. Hao, R.; Hong, L. Compositional coding capsule network with k-means routing for text classification. arXiv 2018, arXiv:1810.09177.
- 62. Vaswani, A.; Shazeer, N.; Parmar, N. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1–15.
- 63. Li, W.; Qi, F.; Tang, M.; Yu, Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* **2020**, *387*, 63–77. [CrossRef]
- 64. Peng, D.L.; Zhang, D.D.; Liu, C. BG-SAC: Entity relationship classification model based on self-attention supported Capsule networks. *Appl. Soft Comput. J.* 2020, *91*, 106–168. [CrossRef]
- 65. Zhao, W.; Chen, X.; Chen, J.; Qu, Y. Sample Generation with Self-Attention Generative Adversarial Adaptation Network (SaGAAN) for Hyperspectral Image Classification. *Remote. Sens.* **2020**, *12*, 843. [CrossRef]