*remote sensing*

MDPI

*Article*

# A Low-Grade Road Extraction Method Using SDG-DenseNet Based on the Fusion of Optical and SAR Images at Decision Level

**Jinglin Zhang [1], Yuxia Li [1,*], Yu Si [1], Bo Peng [1], Fanghong Xiao [1], Shiyu Luo [1] and Lei He [2]**

[1] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2018060507021@std.uestc.edu.cn (J.Z.); 201921060524@std.uestc.edu.cn (Y.S.); 2018060507023@std.uestc.edu.cn (B.P.); xiao_fanghong@std.uestc.edu.cn (F.X.); shiyuluo@uestc.edu.cn (S.L.)
[2] School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China; helei1978@cuit.edu.cn
* Correspondence: liyuxia@uestc.edu.cn

**Abstract:** Low-grade roads have complex features such as geometry, reflection spectrum, and spatial topology in remotely sensing optical images due to the different materials of those roads and also because they are easily obscured by vegetation or buildings, which leads to the low accuracy of low-grade road extraction from remote sensing images. To address this problem, this paper proposes a novel deep learning network referred to as SDG-DenseNet as well as a fusion method of optical and Synthetic Aperture Radar (SAR) data on decision level to extract low-grade roads. On one hand, in order to enlarge the receptive field and ensemble multi-scale features in commonly used deep learning networks, we develop SDG-DenseNet in terms of three modules: stem block, D-Dense block, and GIRM module, in which the Stem block applies two consecutive small-sized convolution kernels instead of the large-sized convolution kernel, the D-Dense block applies three consecutive dilated convolutions after the initial Dense block, and Global Information Recovery Module (GIRM) combines the ideas of dilated convolution and attention mechanism. On the other hand, considering the penetrating capacity and oblique observation of SAR, which can obtain information from those low-grade roads obscured by vegetation or buildings in optical images, we integrate the extracted road result from SAR images into that from optical images at decision level to enhance the extraction accuracy. The experimental result shows that the proposed SDG-DenseNet attains higher *IoU* and *F*1 scores than other network models applied to low-grade road extraction from optical images. Furthermore, it verifies that the decision-level fusion of road binary maps from SAR and optical images can further significantly improve the *F*1, *COR*, and *COM* scores.

**Keywords:** low-grade road extraction; remote sensing; image segmentation; SAR image; deep learning



**Citation:** Zhang, J.; Li, Y.; Si, Y.; Peng, B.; Xiao, F.; Luo, S.; He, L. A Low-Grade Road Extraction Method Using SDG-DenseNet Based on the Fusion of Optical and SAR Images at Decision Level. *Remote Sens.* **2022**, *14*, 2870. https://doi.org/10.3390/rs14122870

Academic Editor: Giuseppe Scarpa

Received: 5 May 2022
Accepted: 14 June 2022
Published: 15 June 2022



## 1. Introduction

Research on extracting road information from remote sensing images has been carried out for many years. However, due to the different width and shape characteristics of different grades of roads, such as national roads, provincial roads, village roads, and mountain roads; roads with different materials have different color and texture characteristics, such as cement, asphalt, earth road, etc.; at the same time, the road area is blocked by buildings, trees, the central green belt of the road and many other factors, so the accurate extraction of road information is still the research frontier and poses a technical difficulty in the field of remote sensing information extraction.

Road extraction can be described as a pixel-level binary classification problem that distinguishes whether each pixel belongs to a road or not [1]. Recently, deep convolution neural networks (DCNNs) have been demonstrated to have significant improvements to typical computer vision tasks such as semantic segmentation [2]. Road semantic segmentation has

applications in many fields, such as autonomous driving [3,4], traffic management [5], and smart city construction [6]. Semantic segmentation requires pixel-level classification [7–9], and it must combine pixel-level accuracy with multi-scale contextual reasoning [7–10]. In general, the simplest way to aggregate multi-scale context is inputting multi-scale information into the network for merging all scales of features. Some researchers have made much progress in the image processing fields. Farabet et al. [11] obtained different scale images by transforming the input image through a Laplacian pyramid. References [12,13] applied multi-scale inputs sequentially from coarse-to-fine. References [7,14,15] directly resized the input image for several scales. Meanwhile, another aggregating multi-scale context way is adopting an encoder-decoder structure, such as SegNet [16], U-Net [17], RefineNet [18], and other networks [19–21], which have demonstrated the effectiveness of models based on encoder-decoder structure. In addition, the context module is an effective way to aggregate multi-scale context information, such as merging DenseCRF [9] into DCNNs [22,23]. The spatial pyramid pool structure is also a common method to aggregate multi-scale context, such as Pyramid Scene Parsing Net (PSP) [24,25].

The larger receptive field is critical for networks because it can capture more global context information from the input images. For a standard convolution neural network (CNN), the traditional way to expand the receptive field is stacking more convolutional layers with a bigger convolutional kernel size, while the operation could result in the exponential expansion of the training parameters, which makes networks hard to train. The alternative way to expand the receptive field is stacking more pooling layers, which can expand the receptive field by reducing the dimension of the feature maps and maintaining the saliency characteristics. Although the pooling operations did not add the training parameters, much information would be lost because of the decrease in spatial resolution.

Reference [23] developed a convolutional network module, dilated convolution, which aggregates multi-scale contextual information without increasing the training parameters and decreasing resolution. Further, the module can also aggregate multi-scale contextual information with different expanding rates of dilated convolution kernel size. Besides, the module can be plugged into existing architectures for any resolution image, which is appropriate for dense prediction. Therefore, DeepLab v2 [26], DeepLab v3 [27], DeepLab v3+ [28], and D-LinkNet [1], which adopted dilated convolution for semantic segmentation, presented better performances.

Another effective strategy to increase the capture capabilities of global features is to introduce attention mechanism. Reference [29] first introduced an attention mechanism into computer vision tasks, which has been proven to be reliable. DANet [30] adopts a spatial and channel attention module to obtain more global context information. CBAM [31] introduced a lightweight spatial and channel attention module. DA-RoadNet [32] constructed a novel attention mechanism module to improve the network's ability to explore and integrate roads.

The network structure for semantic segmentation was divided into several parts, and those networks [1,26–28] only adopted dilated convolutions in one part. In fact, the encoder part and decoder part of existing architectures for semantic segmentation is built by stacking residual blocks or dense blocks. So, dilated convolution layers after each block have been added to capture more global context information. In the research, a new structure, D-Dense blocks, combined with traditional convolution layers and dilated convolution layers, has been proposed. Further, a network is built with D-Dense block and the center part of D-LinkNet for road extraction from satellite images. To increase the capabilities of capturing global features, the DA mechanism [30] is also introduced into the network. With the above design, the dilated convolution can run through the whole network and effectively integrate with the attention mechanism to obtain more global features and information. The presented context network was evaluated through controlled experiments with the Massachusetts Road dataset. The experiments demonstrate that the D-Dense block with attention mechanism architectures reliably increases the pixel-level accuracy for semantic segmentation.

Since SAR has the advantages of all-weather and strong penetration, using SAR images has irreplaceable advantages in remote sensing road extraction, which can further improve the accuracy of road information extraction. Many traditional road segmentation methods of SAR images have been proposed and proved effective. Methods based on human–computer interaction are called semi-automatic methods. Bentabet, L et al. [33] were the first to use the snake model for SAR image road extraction. The results of the experiments show that straight or curved roads could be accurately extracted by this model, but this model needs a large number of human–computer interactions [34]. Some automatic methods were also proven to be useful. Cheng Jianghua et al. [35] proposed a method based on the Markov random field (MRF). In order to maximize calculation efficiency, this method is developed on GPU-accelerated road extraction. Besides, there are also Deep-Learning methods of road extraction on SAR images. Wei X et al. [36] used Ordinal Regression and introduced Road-Topology Loss, which improves the baseline up to 11.98% in the *IoU* metric in their own dataset.

Focused on some problems of low-grade roads in remote sensing images, we study how to improve the accuracy of the road extraction in complex scenes using the powerful feature expression ability of deep learning and the penetrating feature of SAR images.

In this paper, we propose a novel deep learning network model called SDG-DenseNet to improve the accuracy of low-grade road extraction from optical remote sensing images. We fuse the extraction results from the SAR image into that of the optical image at the decision level, which improves the accuracy of low-grade road extraction in practical application scenarios. Therefore, the main contribution of this study can be summarized as:

(1) A novel SDG-DenseNet network for low-grade road extraction in optical images is proposed. The stem block is taken as the starting module to expand the receptive field and preserve image information, while the stem block also reduces the number of parameters. A novel D-dense block is introduced to construct the encoder and decoder of the network, which applies the dilated convolution in all parts from the encoder to the decoder to improve the receptive field of the network. Moreover, in order to make the dilated convolution run through the entire network, this paper introduces a GIRM module combining the dilated convolution and a double self-attention mechanism. The introduction of the GIRM module aims to enhance the network's ability to obtain global information. The segmentation effect of the novel network is better than that of many existing networks;

(2) A decision-level fusion method is proposed for the low-grade road extraction based on optical images and SAR images, which repairs some interrupted roads in the optical image extraction results. The extraction accuracy of decision-level fusion methods is higher than that of optical image-based deep learning methods in practical application scenarios.

## 2. Methods

In order to improve the image semantic segmentation accuracy, a novel SDG-Densenet network for low-grade road extraction in optical images is proposed. The construction of the novel SdG-Densenet for optical image semantic segmentation is composed of three parts: an encoder path, a decoder path, and the center part—the global information recovery module. The encoder path takes RGB images as input parameters and extracts features by stacking convolutional layers and pooling layers. The decoder path restores the detailed information and expands the spatial dimensions of the feature maps with deconvolutional layers. The center part is responsible for enlarging the receptive field, integrating multi-scale features, and maintaining the detailed information simultaneously. The skip connection encourages the reuse of the feature maps to help the decoder path recover spatially detailed information. Besides, a decision-level fusion method is also introduced in order to fuse the results optical image and SAR image, which mainly contains six steps: data preparation, pretreatment, image registration, road extraction, road segmented, and decision level integration. Figure 1 shows the overall structure of the proposed method.

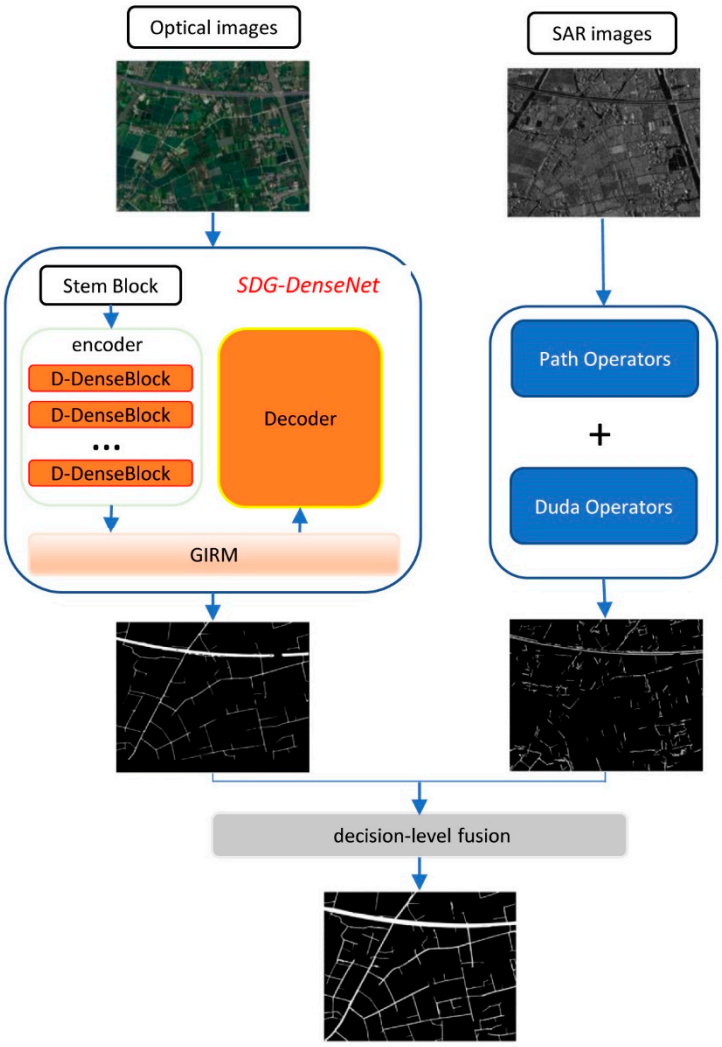



**Figure 1.** The overall technical flow of the proposed method.

### *2.1. Architecture of SDG-DenseNet Network*

Because low-grade roads are easily blocked by vegetation or buildings, there are often problems of fracture and discontinuity in extracting low-grade roads in optical images. At the same time, due to the low construction standard of low-grade roads, their materials are often consistent with the surrounding environment, and they are often integrated into the background in the optical orthographic projection, resulting in a poor extraction effect. Based on the above problems, it is imperative to specialize in the novel network and improve the ability of global information extraction.

Based on D-LinkNet, the SDG-DenseNet was proposed. In order to improve the extraction ability of global information, the global information recovery module was introduced to the proposed new Network for semantic segmentation. Furthermore, the novel network took DenseNet as its backbone instead of ResNet and replaced the initial block with the stem block. Additionally, the Attention mechanism was introduced to improve the ability to obtain global information. The The overall structure of SDG-DenseNet network is shown in Figure 2.

### *2.2. Improved D-Dense Block and Stem Block*

The construction of the D-Dense block is shown in Figure 3. In contrast to the original Dense block, we added three consecutive dilated convolution layers with different expanding rates after the original Dense block. The expanding rates of these three dilated

convolutions are 2, 4, and 8, respectively. The structure of each dilated convolution could be set as BN-ReLU-Conv (1 × 1)-BN-ReLU-D_Conv (3 × 3, rate = 2, or 4, or 8). The same computation process with the original Dense block repeated ($n$ + 3) times and makes the D-Dense block generate feature maps with ($n$ + 3) × $k$ channels.

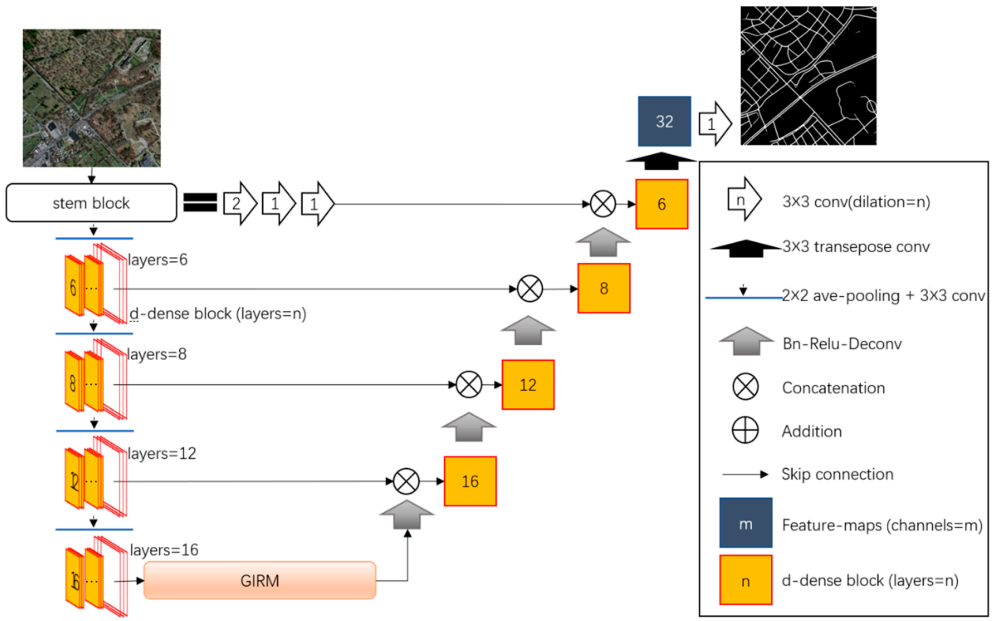



**Figure 2.** The construction of the SDG-DenseNet.

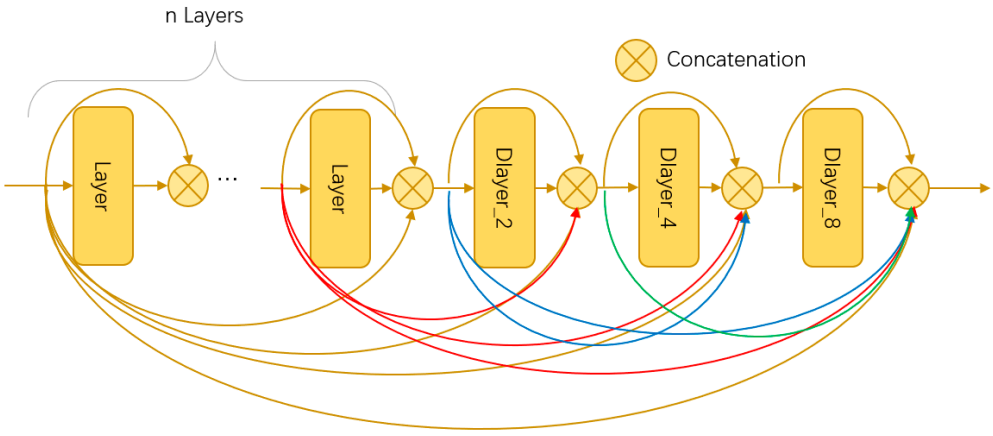



**Figure 3.** The construction of the D-Dense block.

The encoder starts with an initial block and performs convolution on the input image with a kernel of 7 × 7 size and a stride of 2 followed by a 3 × 3 max pooling. In addition, the output channels of the initial block are 64. Inspired by Inception v3 [37] and v4 [38], References [39,40] replaced the initial block [41] 7 × 7 convolution layer, stride = 2 followed by a 3 × 3 max pooling layer by the stem block. The Stem block is composed of three 3 × 3 convolution layers and one 2 × 2 mean pooling layer. The stride of the first convolution layer is 2 and the others are 1. In addition, the output channels for all the three convolution layers are 64. The experiment results in Reference [40] proved that the initial block applied would lose much information due to two consecutive down-sample operations, making it hard to recover the marginal information of the object in the decoder phase. The stem block is helpful for object detection, especially for small objects. So, the research also adopts the stem block at the beginning of the encoder phase.

### *2.3. Global Information Recovery Module (GIRM) Based on d-Blockplus and Attention Mechanism*

In order to weaken and eliminate the problem of road fracture or low recall in low-grade road extraction, this paper proposes a global information recovery module, which is composed of a dual attention mechanism and d-blockplus. The global information extraction module aims to further improve the network's ability to obtain global information to ensure the integrity of the extraction results.

As shown in Figure 4, the global information extraction module is mainly composed of two parts. The dual attention mechanism mainly starts from the two directions of spatial attention and channel attention, extracts and integrates the global information of space and channel, and improves the attention to road targets. d-blockplus introduces multi-layer hole convolution to improve the receptive field, so as to improve the ability of the network to maintain the integrity of road extraction.

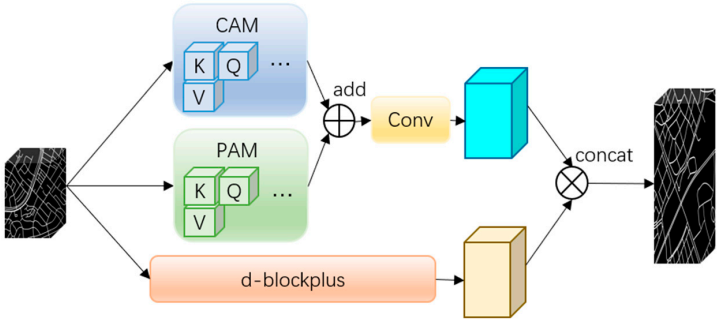



**Figure 4.** The construction of GIRM.

In the center part of the SDG-DenseNet, in addition to the d-blockplus, the position attention module (PAM) and the channel attention module (CAM) are also introduced. PAM and CAM are two reliable self-attention modules, which improve the ability of the network to obtain global information in the spatial dimension and channel dimension, respectively.

Figure 5 shows the structure of PAM. In PAM, the input feature maps go through two branches, and one of them will be used as $Q$ and $K$ to generate a $(H \times W) \times (H \times W)$ Attention probability map. In another branch, it is used as $V$. Where, $V$, $Q$, and $K$ represent value features, query features, and key features, respectively; $C$, $H$, and $W$ represent the channel, height, and weight of the characteristic graph, respectively. The overall structure of PAM is shown in Equation (1):

$$\begin{aligned} Att &= softmax(Q_{(C\times HW)} \cdot K_{(HW\times C)}) \\ F_{out} &= (V_{(C\times HW)} \cdot Att).reshape(C \times H \times W) + Input_{(C\times H\times W)} \end{aligned} \tag{1}$$

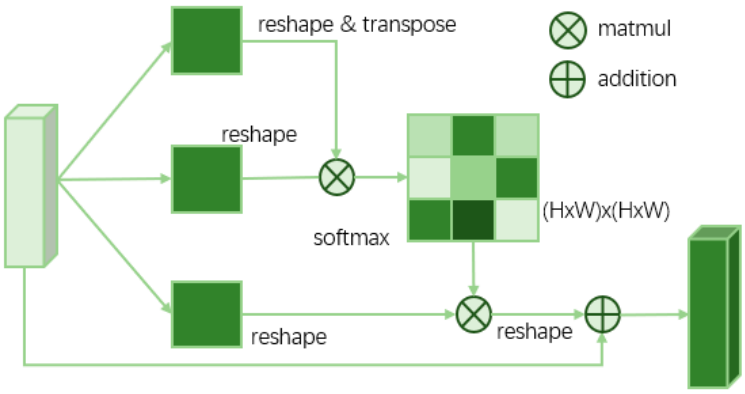



**Figure 5.** Architecture of the position attention module [30].

Figure 6 shows the structure of CAM. The structure of CAM is basically similar to that of PAM. CAM pays more attention to the information on the channel. In this network structure, the size of the probability map generated by CAM is $(C \times C)$, which helps to boost feature discrimination. The overall structure of CAM is shown in Equation (2):

$$\begin{aligned} Att &= softmax(Q_{(C\times HW)} \cdot K_{(HW\times C)}) \\ F_{out} &= (Att \cdot V_{(C\times HW)}).reshape(C \times H \times W) + Input_{(C\times H\times W)} \end{aligned} \tag{2}$$

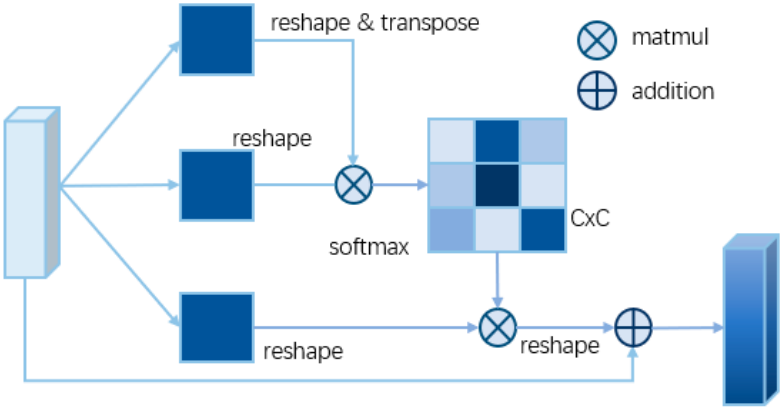



**Figure 6.** Architecture of the channel attention module [30].

D-block has four paths that contain dilated convolution in two cascade modes and two parallel modes, respectively. In each path, dilated convolutions are stacked with different expanding rates. Consequently, the receptive field of each path is different, and the network can aggregate multi-scale context information. Inspired by MobileNetV2 [42], to save network parameters and improve network performance, the bottleneck block is introduced into d-block to build d-blockplus. Figure 7 shows the structure of D-blockplus.

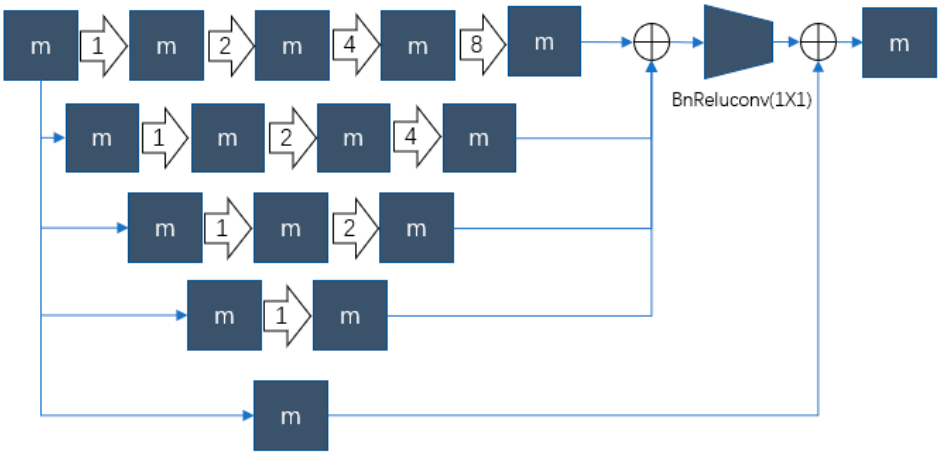



**Figure 7.** The construction of d-blockplus.

### *2.4. Decision-Level Fusion Algorithm for Low Grade Roads*

In optical images, low-grade roads often show the problem where the roads are blocked by buildings, vegetation, shadows, and so on. However, the background of buildings and vegetation is often quite different from the road, and the blocked part is often not judged as a road in the process of deep learning, which directly leads to the phenomenon of fracture or undetected in the extraction results of low-grade roads. Figure 8 shows some examples of blocked roads. In these pictures, the roads in red boxes show fractures in the optical image because it is obscured by vegetation, buildings, or shadows.

**Figure 8.** Figures of blocked roads.

For the problems of the above complex scenes, the imaging mechanism of the optical image determines that the SDG-DenseNet network model cannot solve the problem of poor road continuity well. In this paper, the optical image extraction results based on the SDG-DenseNet network model and the SAR image extraction results based on Duda and path operators [43] realize decision-level fusion.

The Duda operator is a linear feature extraction operator that divides an $N \times N$ window into three parallel linear parts. The specific structure of the Duda operator is shown in Figure 9, where A, B, C, C1, and C2 represent the mean gray values of the three parts. What's more, the operator shown in Figure 9a has a relatively strong ability to extract roads in the horizontal direction, and the operator shown in Figure 9b has a relatively strong ability to extract roads with a certain inclination angle.

The other two types of Duda Operators are a 90-degree rotation of the above two. The function to determine the new value of a pixel can be expressed as follows:

$$H(x) = (1 - \frac{C}{A})(1 - \frac{C}{B})\frac{C1}{C2}. \quad (3)$$

Path operators refer to path openings and closings, which are morphological filters applied to analyze oriented linear structures in images. The morphological filter defines the adjacency graphs as structuring elements. Four different adjacency graphs are defined as horizontal lines, vertical lines, and two diagonal lines, respectively. Applying these four adjacency graphs to a binary image, the maximum path length of each pixel can be achieved. Then, the pixels, whose maximum path lengths are larger than the threshold Lmin, are retained in the image.

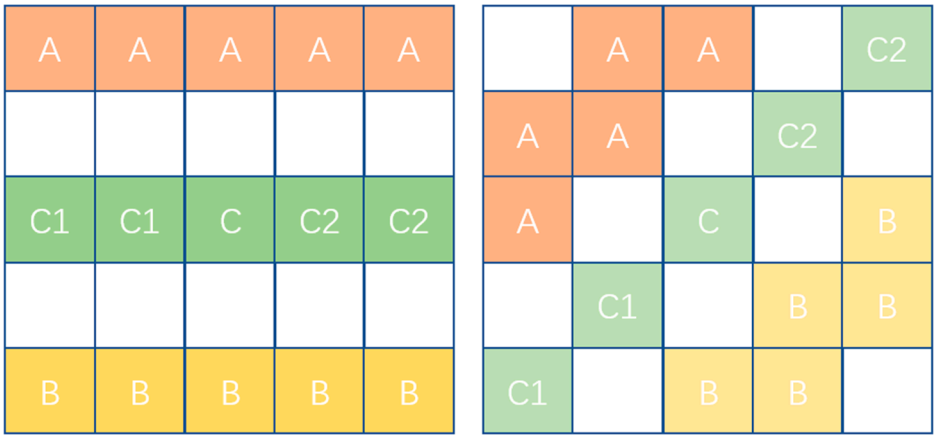



**Figure 9.** Window structure of two types of Duda operators.

The specific algorithm flow of the decision-level fusion method for low grade roads is shown in Figure 10.

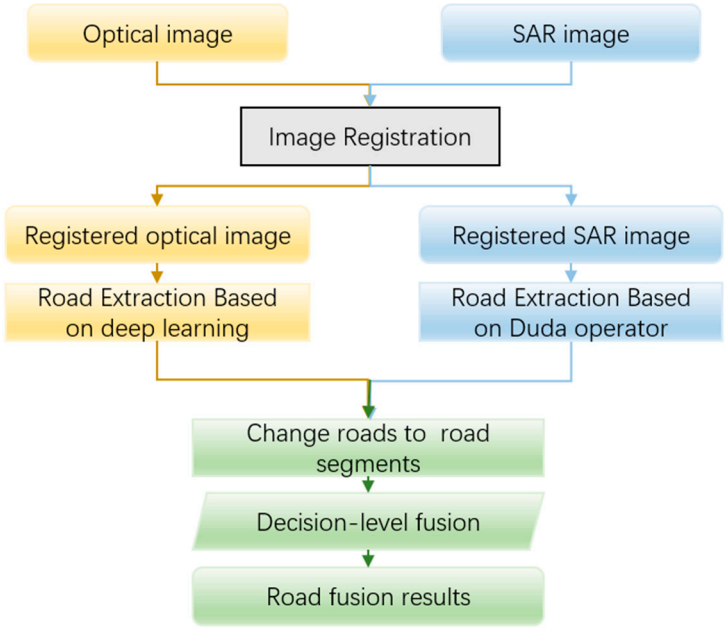



**Figure 10.** Low-grade road extraction algorithm for decision-level fusion of optical and SAR images.

Figure 10 shows the overall technical process of the road extraction algorithm based on the decision-level fusion of high-resolution optical and SAR remote sensing images. The specific steps of the algorithm are as follows.

***Step 1***: Data preparation. Obtain optical remote sensing images and SAR images in the same area, and their imaging time should be as close as possible;

***Step 2***: Pretreatment. The optical remote sensing image and SAR image are preprocessed, respectively, including radiometric correction, geometric correction, geocoding, and so on;

***Step 3***: Image registration. The optical remote sensing image and SAR image are matched and transformed into the same pixel coordinate system;

***Step 4***: Road extraction. Roads in optical remote sensing images are extracted by SDG-DenseNet and those in SAR images are extracted by the method in Reference [43], which is based on Duda and Path operator;

***Step 5***: Roads segmented. For the road extraction results of optical remote sensing image and SAR image, the road segments are obtained by segment method, and the attributes of each segment are recorded;

***Step 6***: Decision level fusion. Taking the line segment as the basic unit, the final road distribution map is obtained by decision-making level fusion of the roads extracted from the optical remote sensing image and SAR image.

## 3. Experiments

Our network experiments are performed on the Massachusetts Roads Dataset, and we test the fusion method in our own dataset that came from WorldView-2, WorldView-4, and TerraSAR-X. The TensorFlow platform was selected as the deep learning framework to train and test all networks. All models are trained on one NVIDIA GTX 2080 Ti GPU.

### *3.1. Dataset and Data Augmentation*

Three sets of satellite images were applied to evaluate the Low-Grade road extraction method. To verify the effectiveness of the proposed SDG-DenseNet network on public datasets, we tested the SDG-DenseNet on the Massachusetts dataset. In addition, we conducted low-grade road extraction experiments on the self-built Chongzhou–Wuzhen dataset. Finally, we conducted decision-level fusion experiments on two sets of large-scale images from the Chongzhou and Wuzhen regions including optical and SAR images.

We trained and tested our SDG-DenseNet network model on the Massachusetts Roads Dataset [44], which consists of 1108 training images, 14 validation images, and 49 test images. The size of each image is 1500 × 1500. We cut each 1500 × 1500 image into four 1024 × 1024 images. Therefore, we obtained 4432 training images, 56 validation images, and 196 test images. Further, we performed data augmentation on the training set, including rotation, flipping, cropping, and color jittering, which could prevent the training set from overfitting. After data augmentation, we obtained 22,160 training images in total. Finally, we obtained 22,160 training images, 56 validation images, and 196 test images.

In order to test the proposed SDG-DenseNet network of low-grade road extraction, this paper also tests the SDG-DenseNet on the self-built dataset: The Chongzhou–Wuzhen dataset. Table 1 displays the three source images of the self-built dataset. We cut the three source images into 13,004 512 × 512 images. Therefore, we obtained 11,788 training images, 204 validation images, and 1012 test images. After the data augmentation of the training set, we got 47,152 training images. Finally, we obtained 47,152 training images, 204 validation images, and 1012 test images.

**Table 1.** The three source images of the self-built low-grade road dataset.

| Data | Satellites | Resolution Ratio | Date | Scale | Area |
|---|---|---|---|---|---|
| 1 | WorldView-4 | 0.6 m | 13 May 2018 (optical) | 3469 × 4786 | Chongzhou, Sichuan |
| 2 | WorldView-2 | 0.5 m | 27 July 2018 (optical) | 2800 × 3597 | Wuzhen, Zhejiang |
| 3 | WorldView-2 | 0.5 m | 27 July 2018 (optical) | 2800 × 1798 | Wuzhen, Zhejiang |

We also test our decision-level fusion experiments on two sets of large-scale images from the Chongzhou and Wuzhen regions including optical and SAR images. The optical images came from WorldView-2 and WorldView-4, while we got the SAR images from TerraSAR-X. As shown in Table 2, in order to test the effect under application conditions, the decision-level fusion experiment is mainly tested on the two large-scale images.

**Table 2.** Two sets of large-scale images used in decision-level fusion experiments.

| Data | Satellites | Resolution Ratio | Date | Scale | Area |
|---|---|---|---|---|---|
| 1 | WorldView-4 | 0.6 m | 13 May 2018 (optical) | 3469 × 4786 | Chongzhou, Sichuan |
| | TerraSAR-X | 0.8 m | 20 September 2018 (SAR) | | |
| 2 | WorldView-2 | 0.5 m | 27 July 2018 (optical) | 2800 × 3597 | Wuzhen, Zhejiang |
| | TerraSAR-X | 0.9 m | 24 March 2019 (SAR) | | |

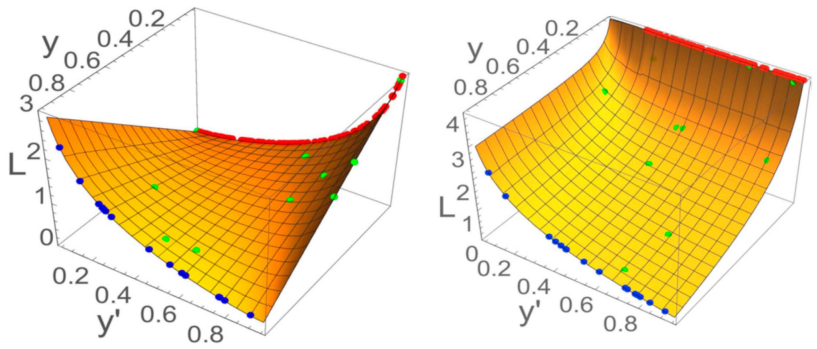
### *3.2. Hybrid Loss Function and Implementation Details*

In previous work, most networks train their models only by using the cross-entropy loss [45], which is defined as Equation (3):

$$L_{ce} = -\frac{1}{N}\sum_{i=0}^{N}(y \log y' + (1-y)\log(1-y')), \tag{4}$$

where $N$ indicates categories. $y$ and $y'$ mean the label and prediction vectors, respectively. Since an image consists of pixels, for road area segmentation, the imbalance of sample points (where the roads only cover a small part of the whole image) makes the direction of the gradient decrease toward the back corner (Figure 11a), which leads to a local optimum, especially in the early stage [46]. The Jaccard loss function is defined as:

$$L_{jaccard} = \frac{1}{N}\sum_{i=0}^{N}\frac{y_i y'_i}{y_i + y'_i - y_i y'_i} \tag{5}$$




(**a**) (**b**)

**Figure 11.** Different loss function surface. (**a**) Cross entropy surface; (**b**) Jaccard surface.

Its surface is shown in Figure 11b. As we can see, the Jaccard loss can address this problem if we sum the Jaccard loss and the cross-entropy loss together. So, the whole loss function is defined as:

$$L = L_{ce} - \lambda \log L_{jaccard}, \tag{6}$$

where $\lambda$ is the weight of the Jaccard loss in the whole loss. Furthermore, the red, green, and blue points in Figure 11 represent the local maxima, saddle points, and local minima on the loss surface, respectively.

In the training phase, we chose Adam as our optimizer and originally set the learning rate to be 0.0001. We reduce the learning rate by 10 times while observing the loss value decreasing slowly. The loss weight $\lambda$ is set to 1. The batch size during the training phase is set to 1.

### *3.3. Decision-Level Fusion Experiment*

To verify the effect of every step in the decision-level fusion method for low-grade roads, we apply the fusion method to the road extraction results from the network and method based on the Duda operator and Path operator, using the large-scale images mentioned in Table 2 and the details of ***Step 6***, where decision-level fusion is operated as in Figure 12. The detailed workflow of the Decision level fusion.

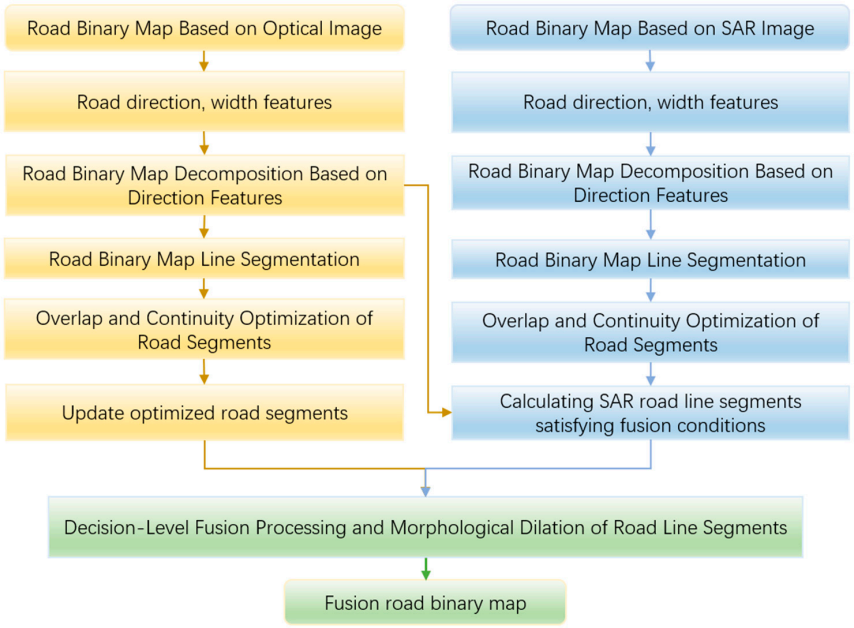



**Figure 12.** The detailed workflow of the decision level fusion.

As shown in Figure 12, the main process is divided into five steps:

***Step 1***: Road binary map extracted from input optical image and SAR image (not segmented);

***Step 2***: Segment the road binary map extracted from the SAR image, including extracting the road feature direction map, decomposing the binary map according to the direction feature, thinning the decomposed layer based on the curve fitting algorithm, and optimizing the line segment overlap, continuity and intersection to obtain the road segment set extracted from the SAR image;

***Step 3***: Segment the road binary map extracted from the optical image, optimize the overlap and continuity of segments, and record the updated segments of continuity optimization;

***Step 4***: For each road segment extracted from the SAR image, we judge whether it meets the fusion conditions with optical image road extraction results according to the overlap ratio in the corresponding optical extraction road binary layer, and record the qualified SAR road segments;

***Step 5***: After morphological expansion according to the width feature, the continuously optimized and updated line segments and the SAR road line segments meeting the fusion conditions are calculated with the original optically extracted road binary map according to pixels to obtain the fused Road Distribution binary map.

The specific method of searching line segments satisfying fusion conditions is shown in Figure 13. Assuming that $A_m$ represents the road area on layer m after the decomposition of optical image extraction results, $L_{mn}$ is the line segment n on layer m from SAR image road extraction results. They belong to the same layer m, that is, the road has similar directional features. We then count the number of pixels $l_{n1}$ and $l_{n2}$ where $L_{mn}$ falls inside and outside the $A_m$ region, and calculate the overlap rate $r = l_{n1}/(l_{n1} + l_{n2})$. If r is greater than the threshold $T_r$, $L_{mn}$ is recorded as the road segment meeting the fusion conditions. In a practical application, the threshold tr takes an empirical value of 0.3. We traverse all

SAR extracted road segments until all SAR image-extracted road segments meeting the above fusion conditions are recorded.

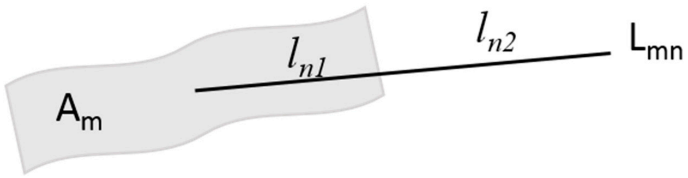



**Figure 13.** Schematic diagram of road fusion condition judgment for optical and SAR extraction.

### *3.4. Evaluation Metrics*

In order to evaluate the performance of different road segmentation models, four evaluation metrics are used to evaluate the extraction results, including intersection-over-union (*IoU*), completeness (*COM*), correctness (*COR*), and *F*1-score [47], which are defined as:

$$IoU = \frac{TP}{TP+FN+FP} \qquad COR = \frac{TP}{TP+FP}$$
$$COM = \frac{TP}{TP+FN} \quad .... \quad F1 = \frac{2\times COM\times COR}{COM+COR} \tag{7}$$

*TP* (True Positive) indicates that the extraction result is determined as a road, which is actually part of the road; *FP* (False Positive) indicates that the extraction result is determined as a road, but it is not actually part of the road; *FN* (False Negative) indicates that the extraction result is determined to be not a road, but it is actually part of the road. The *COM* scores of different models show the ability to maintain the completeness of the segmented roads. The higher the score, the better the road continuity extracted by the model. The *COR* scores of different models show the ability on reducing false detection of the segmented roads. The higher the score, the fewer areas will be falsely detected. The *IoU* and *F*1 scores are the overall evaluation metrics that synthesize *COM* and *COR* scores and evaluate the overall quality of segmentation results.

Based on these evaluation metrics, we can obtain the performance of model road extraction results in different aspects from *COM* and *COR* scores, and obtain the overall performance judgment from *F*1 and *IoU* scores.

## 4. Results and Discussion

### *4.1. Results of the Massachusetts Roads Dataset*

In order to further verify the effectiveness of the proposed method, we evaluated our network with Massachusetts Roads Dataset. We divided the test images into two levels—general scene and complicated scene—according to the complexity of the image content scene. The sample results are shown in Figures 14–18.

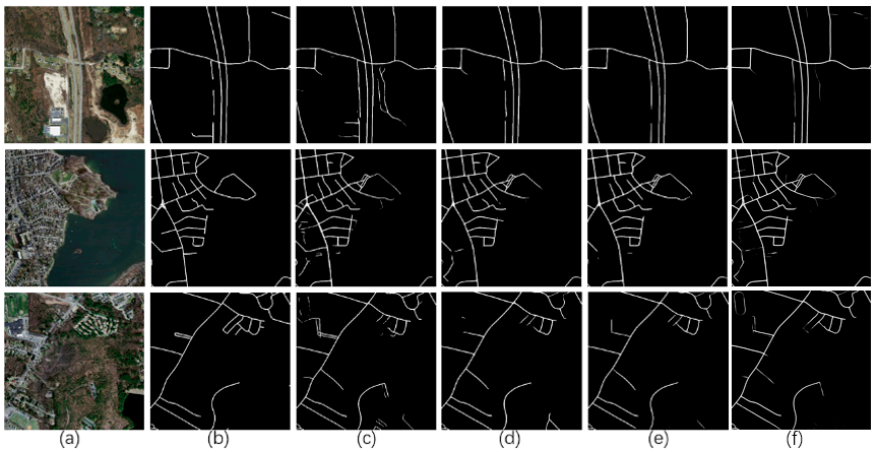



**Figure 14.** Road extraction results in general scene images; (**a**) input image; (**b**) label image; (**c**) D-LinkNet; (**d**) S-DenseNet; (**e**) SD-DenseNet; (**f**) SDG-DenseNet.

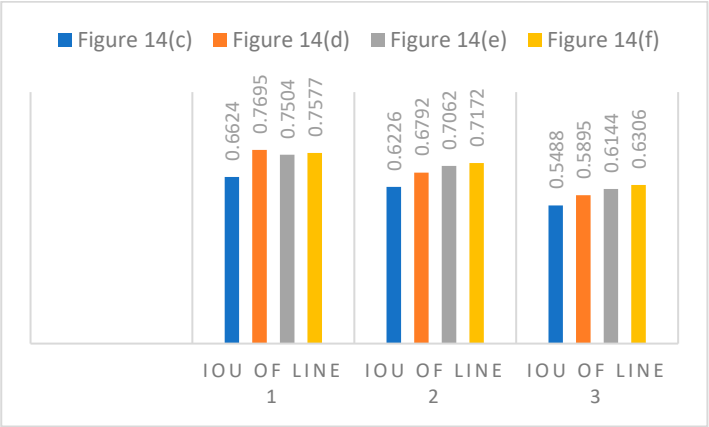



**Figure 15.** *IoU* scores of the methods in Figure 14.

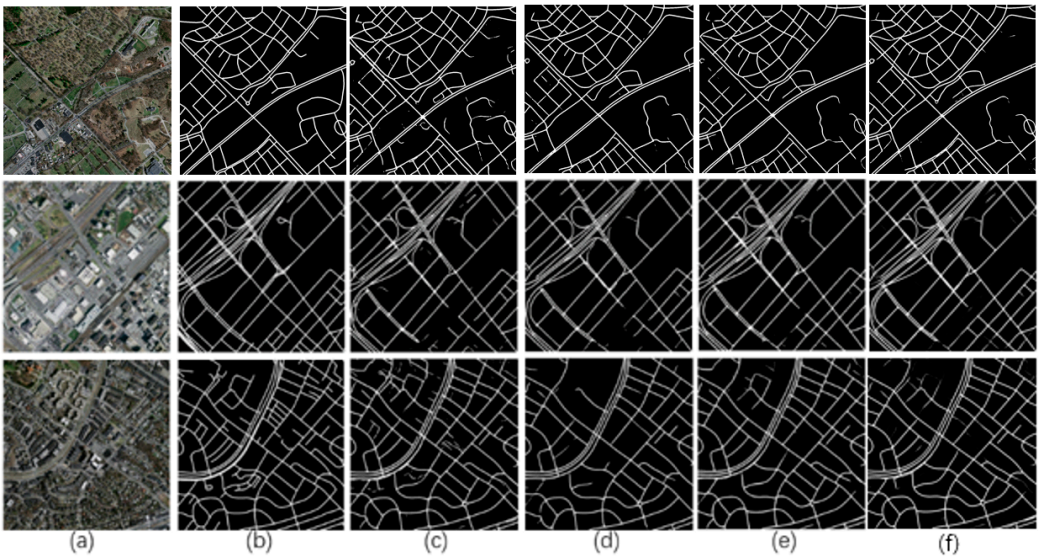



**Figure 16.** Road extraction results in complicated scene images (**a**) input image; (**b**) label image; (**c**) D-linkNet; (**d**) S-DenseNet; (**e**) SD-DenseNet; (**f**) SDG-DenseNet.

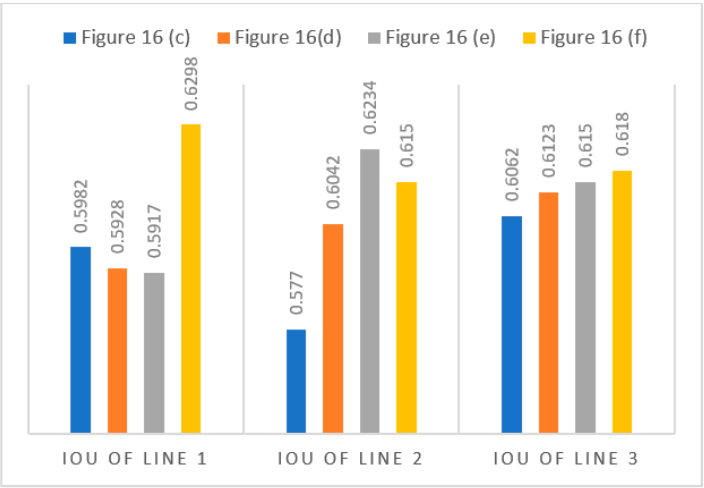



**Figure 17.** *IoU* scores of the methods in Figure 16.

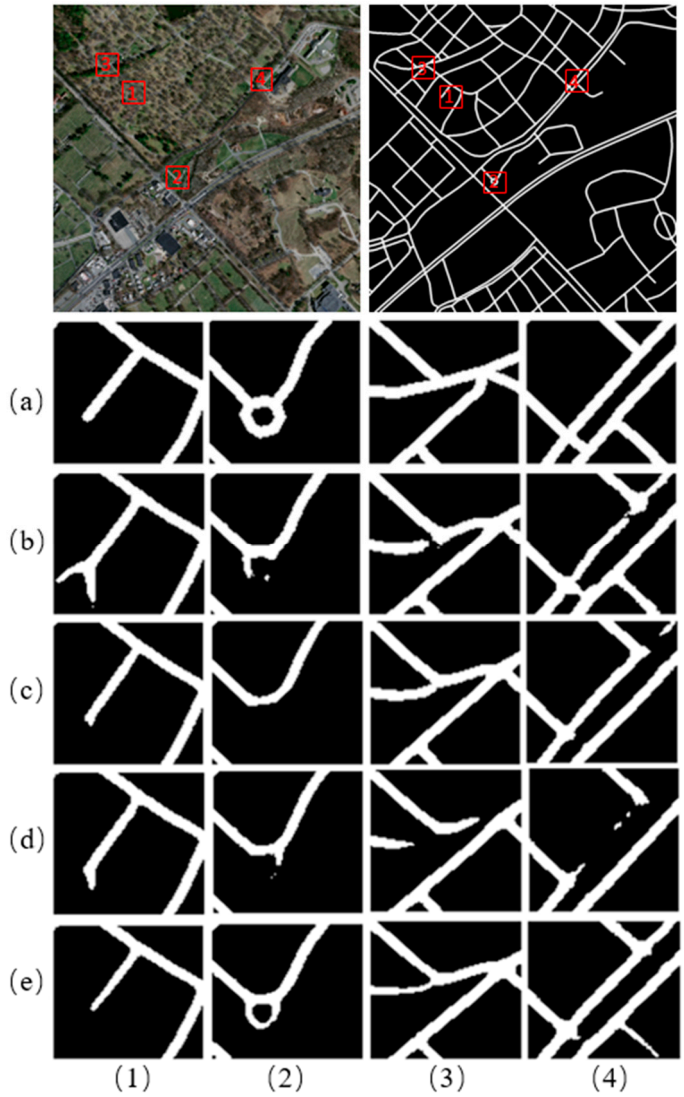



**Figure 18.** Four detailed areas of road extraction results in complicated scene images (**a**) area of label image; (**b**) D-LinkNet; (**c**) S-DenseNet; (**d**) SD-DenseNet; (**e**) SDG-DenseNet.

Figures 14 and 15 show the extraction results of general scene images. D-LinkNet shows the network built on residual blocks and the encoder part, DenseNet shows the network built on the Dense block, S-DenseNet shows the network built on Dense block and Stem block, SD-DenseNet represents that the network has also added dilated convolution on the basis of the previous networks, and SDG-DenseNet is built on the basis of the Stem block, D-Dense Blocks, and the GIRM module. The extraction results of the DLinkNet model contain some redundant information, and many independent patches are left in the image, which could affect the result of the overall accuracy. The parking lot areas, which are similar to roads, were successfully identified as backgrounds. However, some roads are not completely extracted. SDG-DenseNet has been further improved to make the completeness of roads better. The information extracted by the SDG-DenseNet network structure is more accurate.

Figure 15 shows the *IoU* scores for each image in each row in Figure 14. The proposed SDG-DenseNet achieves high *IoU* scores under all three optical images, which are 9.53%, 9.46%, and 8.18% higher than the baseline D-LinkNet, respectively.

Figure 16 shows the three extraction results of the complicated scene images from the 49 test images. Each road network includes more different level roads and flyover roads.

These complex situations seriously affect the road extraction results of every network model. However, the SDG-DenseNet can better extract every road including shadow obscured roads.

Figure 17 shows the *IoU* scores for each image in each row in Figure 16. Similar to Figure 15, in the three optical images, the proposed SDG-DenseNet achieves high *IoU* scores, which are 3.16%, 3.8%, and 1.18% higher than the baseline, respectively. At the same time, in order to improve the reliability of the results from a methodological point of view, the effects of different modules on different images are often different, for example, the *IoU* of Line 1 in Figure 17 shows that the two improved methods perform less well on this optical image. However, it is worth mentioning that the comprehensive results of statistics show that the average *IoU* score of the optimized model on the test set (196 test images) is higher than that of the baseline.

Figure 18 shows some detailed areas of the first image in Figure 16. The different results on Area 1 and Area 2 implied that SDG-DenseNet has a better ability on the correctness of segmentation. Area 1 shows the novel network's improvement in avoiding false segmentation, while in Area 2 it also emerges that SDG-DenseNet performs well in the recall ratio of extraction. Besides, Area 3 and Area 4 show that the novel network also performs perfectly when focusing on the completeness of the result of road extraction. In Figure 18, column (1), column (2), column (3), and column (4) correspond to area 1, area 2, area 3, and area 4 respectively.

Figures 14–18 show the semantic segmentation results of some randomly selected images in the Massachusetts test set. In order to further prove the effectiveness of the improved model on the test data set, this paper counts the evaluation indicators of the segmentation results of different models on the test set (196 test images). Through the training model and experiment, we get the D-LinkNet, DenseNet, S-DenseNet, SD-DenseNet, and SDG-DenseNet evaluation metrics index, as shown in Table 3. We found the *IoU* and *F*1 scores of the network built on the Dense block or D-Dense block to be much higher than the network built with the residual block. Besides, the model based on DenseNet with the D-Dense block has higher *IoU* and *F*1-scores than that with Dense block.

**Table 3.** Results of the Massachusetts Roads Dataset of different models. The bold font indicates the optimal value under the current evaluation metrics.

| Model's Description | *F*1 | *IoU* | *COR* | *COM* |
|---|---|---|---|---|
| D-LinkNet | 0.7688 | 0.6286 | 0.7712 | 0.7727 |
| DenseNet | 0.7786 | 0.6423 | 0.7780 | **0.7854** |
| S-DenseNet | 0.7810 | 0.6462 | 0.8153 | 0.7557 |
| SD-DenseNet | 0.7894 | 0.6562 | **0.8190** | 0.7667 |
| SDG-DenseNet | **0.7963** | **0.6647** | 0.8186 | 0.7767 |

In other words, compared with D-LinkNet, the novel network can extract roads more correctly and maintain good road completeness. Furthermore, when comparing the stem block with the initial block, we find that the network with the stem block is much better than the initial block in the correctness of road extraction. At the same time, stem block also improves the *IoU* and *F*1 scores. The experiment results show the SDG-DenseNet could obtain better *IoU* and *F*1 scores when performing well in the correctness of road extraction. It can also be seen from the table that the SDG-DenseNet is more balanced than other networks in its ability to maintain road completeness and correctness, while both *COR* and *COM* indices are kept at a relatively high level, thus achieving higher *F*1 and *IoU* Scores.

### *4.2. Results on Massachusetts Roads Dataset of Different Methods*

To evaluate our method performance, we compare its *IoU* scores with Residual Unet [46], Joint-Net [48], Dual Path Morph-Unet [49], and DA-RoadNet [32] which have been used in road extraction from satellite images.

Table 4 shows the scores obtained by different methods on the Massachusetts Roads Dataset. The SDG-DenseNet had the highest *F*1 and *IoU* which proves the excellent performance of SDG-DenseNet in road extraction. Besides, as shown in Table 4, our new network achiever a higher *COM* score than other networks, while the *COR* score of the SDG-DenseNet is not much lower than other networks. In other words, our network achieves a good balance in maintaining the completeness and correctness of segmentation.

**Table 4.** Results of the Massachusetts Roads Dataset of different methods. The bold font indicates the optimal value under the current evaluation metrics.

| Method | *F*1 | *IoU* | *COR* | *COM* |
|---|---|---|---|---|
| Residual Unet [46] | * | 0.6340 | * | * |
| Joint-Net [48] | 0.7805 | 0.6310 | **0.8536** | 0.7190 |
| Dual Path Morph-Unet [49] | * | 0.6440 | * | * |
| DA-RoadNet [32] | 0.7819 | 0.6419 | 0.8524 | 0.7124 |
| SDG-DenseNet (ours) | **0.7963** | **0.6647** | 0.8186 | **0.7767** |

('*' represents the metrics not mentioned in the cited papers).

### *4.3. Results of Low-Grade Roads on the Chongzhou–Wuzhen Dataset*

In order to fully uncover the characteristics of the low-grade road extraction task and the performance of different networks on this task, this paper tests the low-grade road in the Chongzhou–Wuzhen dataset. In the test process, according to whether the low-grade roads are blocked, the complexity of the low-grade road structure, and the complexity of the background scene, the extraction difficulty is divided into four cases: simple, general, and complicated.

Figures 19–21 shows the extraction results of four different network models in simple scenes, general scenes, and complex scenes, respectively. Figure 19 shows the detection result in a simple scenario. The detection effect of D-LinkNet in a simple scenario is best, especially for the detection of the expressway; while its integrity is higher, there are fewer false detection parts. Figure 20 shows the detection effect in a general scenario. At this time, D-LinkNet has obvious road fracture and missing detection. SDG-DenseNet has the best detection effect. Compared with other networks, it extracts the most complete roads. Figure 21 shows the detection effect in complex scenes, and several networks show different degrees of missed detection and false detection. S-DenseNet shows the strongest ability to maintain integrity, but there are many false detection areas; SDG-DenseNet has a certain degree of road fracture, but there are few false detections.

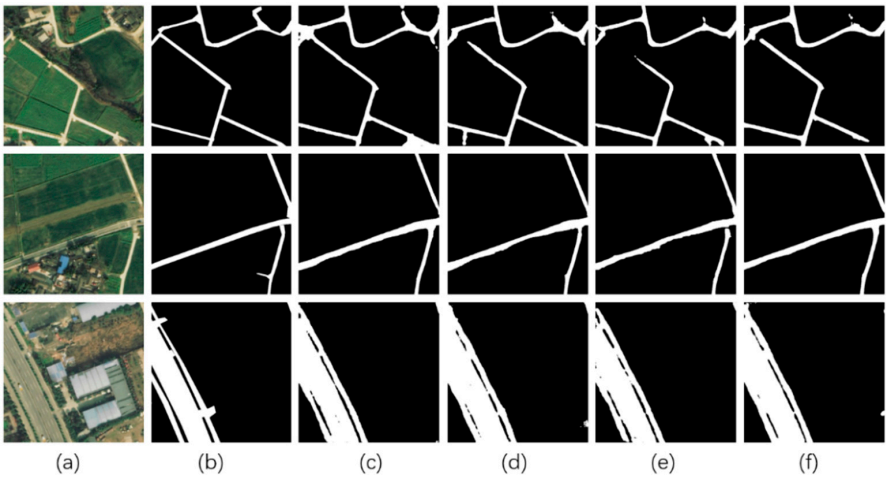



**Figure 19.** Low-grade road extraction in simple scene images (**a**) input image; (**b**) label image; (**c**) D-LinkNet; (**d**) S-DenseNet; (**e**) SD-DenseNet; (**f**) SDG-DenseNet.

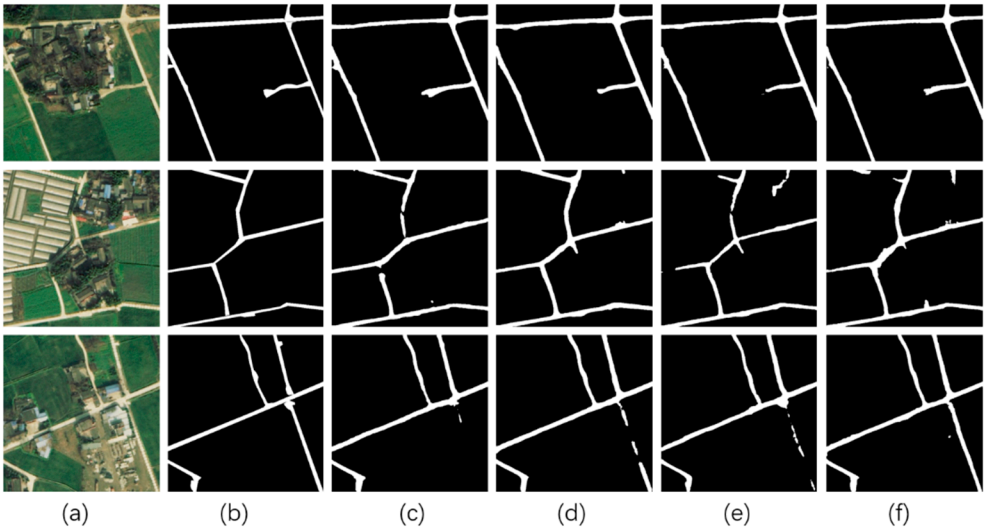



**Figure 20.** Low-grade road extraction in general scene images (**a**) input image; (**b**) label image; (**c**) D-LinkNet; (**d**) S-DenseNet; (**e**) SD-DenseNet; (**f**) SDG-DenseNet.

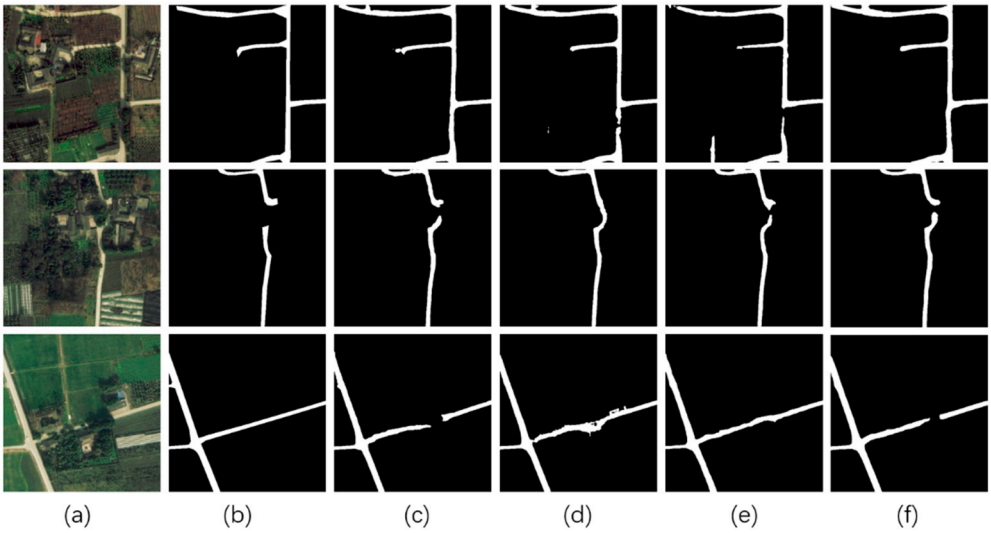



**Figure 21.** Low-grade road extraction in complicated scene images (**a**) input image; (**b**) label image; (**c**) D-linkNet; (**d**) S-DenseNet; (**e**) SD-DenseNet; (**f**) SDG-DenseNet.

Table 5 shows the *IoU* scores of the different models on the test of the Chongzhou–Wuzhen dataset. The result shows that SDG-DenseNet achieved the highest *IoU* scores while its model size is much less than D-LinkNet, which proves that the SDG-DenseNet has the best performance on low-grade road extraction tasks. S-DenseNet has the least parameters of the four networks, which is mainly due to the reduction of parameters by dense block.

**Table 5.** Results of the Chongzhou–Wuzhen test set on different models. The bold font indicates the optimal value under the current evaluation metrics.

| Model's Description | *IoU* | Model Size |
|---|---|---|
| D-LinkNet | 0.5236 | 0.98 GB |
| S-DenseNet | 0.5558 | **81.7 MB** |
| SD-DenseNet | 0.5796 | 106 MB |
| SDG-DenseNet | **0.5901** | 265 MB |

### *4.4. Extraction Results of Low-Grade Roads on Large-Scale Images of the Fusion Method*

In order to verify the feasibility and effect of the decision-level fusion method, and to test the overall effect of the process in the actual application scenario, we extracted the optical image based on SDG-DenseNet and the SAR image based on the Duda operator for the two large-scale images mentioned in Table 2, and then tested the effect of decision-level fusion.

In order to more intuitively reflect the effect of the fusion method, we compare the extracted roads with the roads in the label.

Figures 22 and 23 show the result of the decision-level fusion method tested on our own dataset in the Chongzhou area.

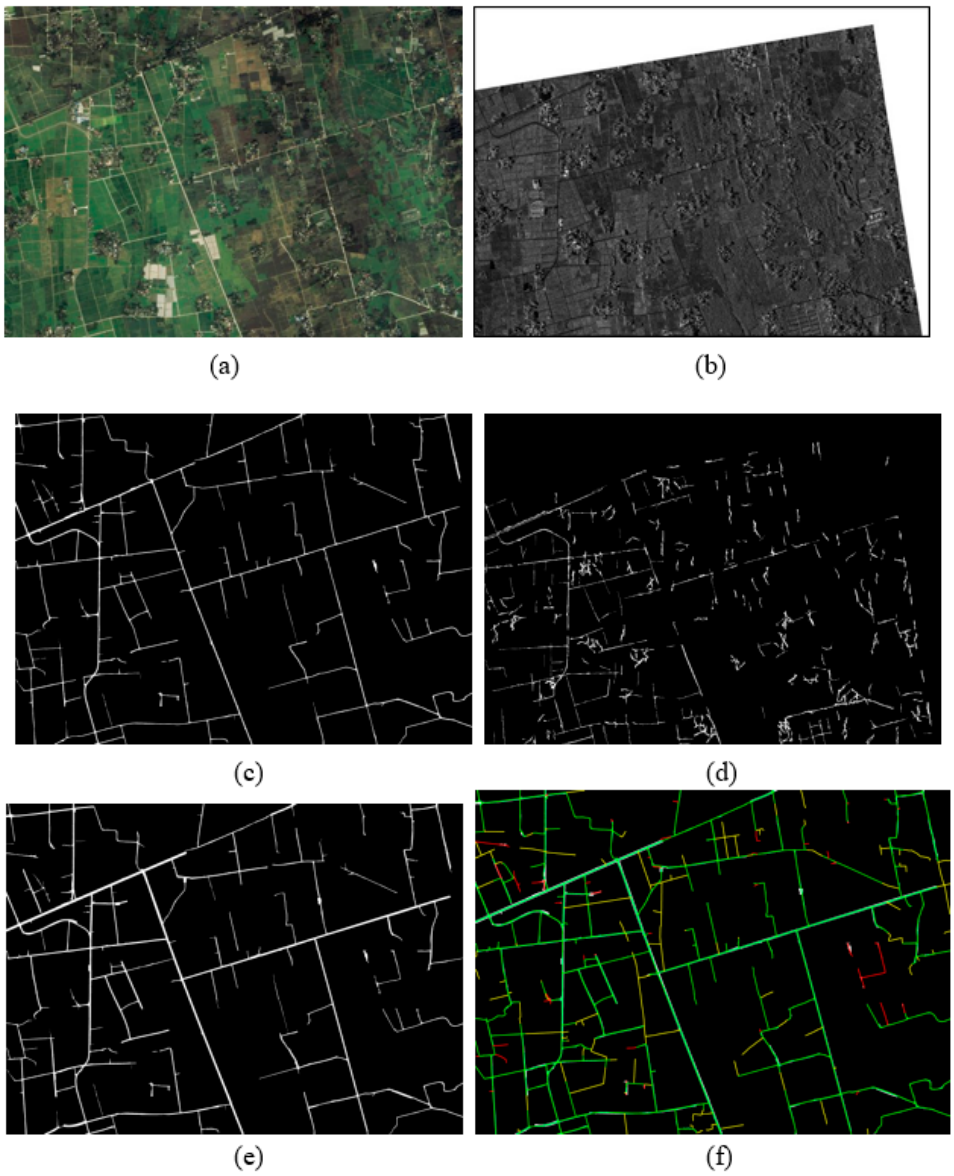



**Figure 22.** Tested data Area 1. Optical and SAR remote sensing images and road extraction results in the Chongzhou area. (**a**) Worldview-4 optical remote sensing image; (**b**) TerraSAR-X remote sensing image; (**c**) road extraction results of optical remote sensing image; (**d**) road extraction results from SAR remote sensing images; (**e**) road fusion extraction results of optical and SAR images; (**f**) ground truth and marking results of road fusion extraction results (green refers to the correctly extracted road, red refers to the incorrectly extracted road, and yellow refers to the omitted real road).

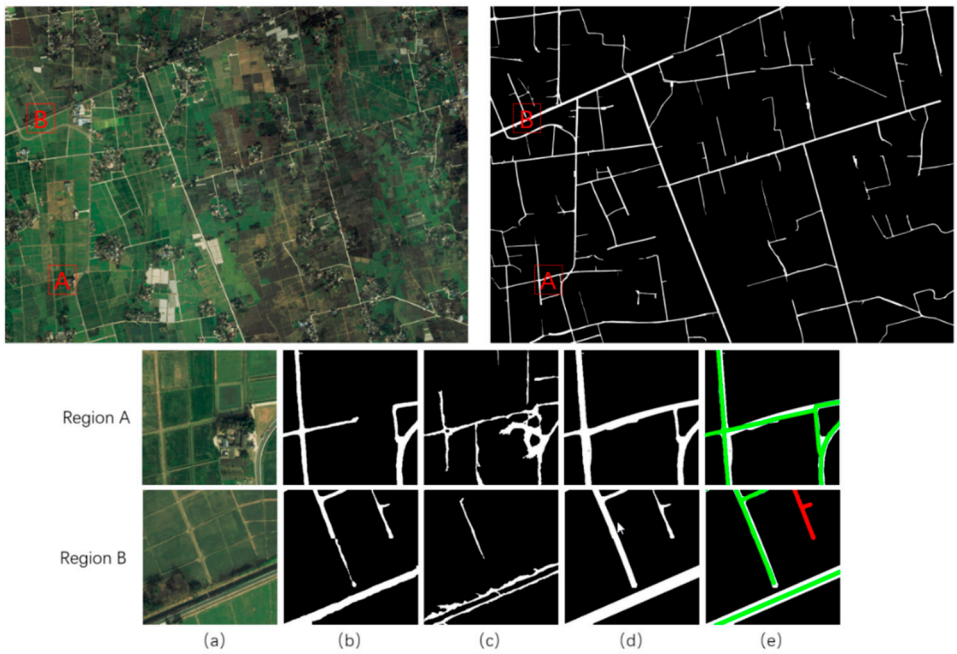



**Figure 23.** Tested data Area 1. Some details in optical and SAR remote sensing images and road extraction results in the Chongzhou area. (**a**) Worldview-4 optical remote sensing image; (**b**) road extraction results of optical remote sensing image; (**c**) road extraction results from the SAR remote sensing images; (**d**) road fusion extraction results of optical and SAR images; (**e**) ground truth and marking results of road fusion extraction results (green refers to the correctly extracted road, red refers to the incorrectly extracted road, and yellow refers to the omitted real road).

Figure 22 display the extraction effect of the optical image, SAR image, and decision-level fusion on low-grade roads in practical application scenarios. The road extracted from the optical image using SDG-DenseNet is more complete and continuous than the road extracted from the SAR image using the Duda and Path operators. However, the extraction results of the SAR images contain some information that are not found in optical image extraction results, such as some roads obscured by vegetation or buildings.

Figure 23 show some details from Figure 22. As shown in the figures, according to the extraction results of SAR images, decision-level fusion fixes some problems of road fracture and missing detection caused by vegetation or building occlusion in optical images.

We also tested the decision-level fusion method in the Wuzhen area of the self-made dataset, as shown in Figures 24 and 25. Figure 24 shows the extraction results in large-size images in practical application scenarios, which are similar to the results in Chongzhou. The extraction results of optical images are good in continuity, but there are also obvious problems of broken road extraction results and missing detection. After fusion with the SAR image, the partially occluded roads become continuous, and some roads that were missed in the optical image were detected.

Figure 25 shows some details of the detection results in Wuzhen. Figure 25 region A(b) shows the extracted road breaks due to bridge interference in the optical image, which can be seen in region A(d). After decision-level fusion, the broken extraction result is fixed. A similar situation occurs in region B due to the occlusion of vegetation, and the roads that are missed in the optical image are also repaired by the decision-level fusion method.

Table 6 shows the results of the two large-scale images mentioned in Table 2. We use the manual interpretation annotation method to evaluate and analyze the low-grade road extraction results, in other words, the matching degree between the extracted road network and the reference road is evaluated through completeness, correctness, and accuracy. As shown in Figures 22–25 and Table 6 through the decision-level fusion extraction of optical and SAR images, the *F*1-scores of road extraction can reach more than 0.85. The *F*1, *COM*,

and *COR* scores are significantly higher than the results using only the optical image extraction method.

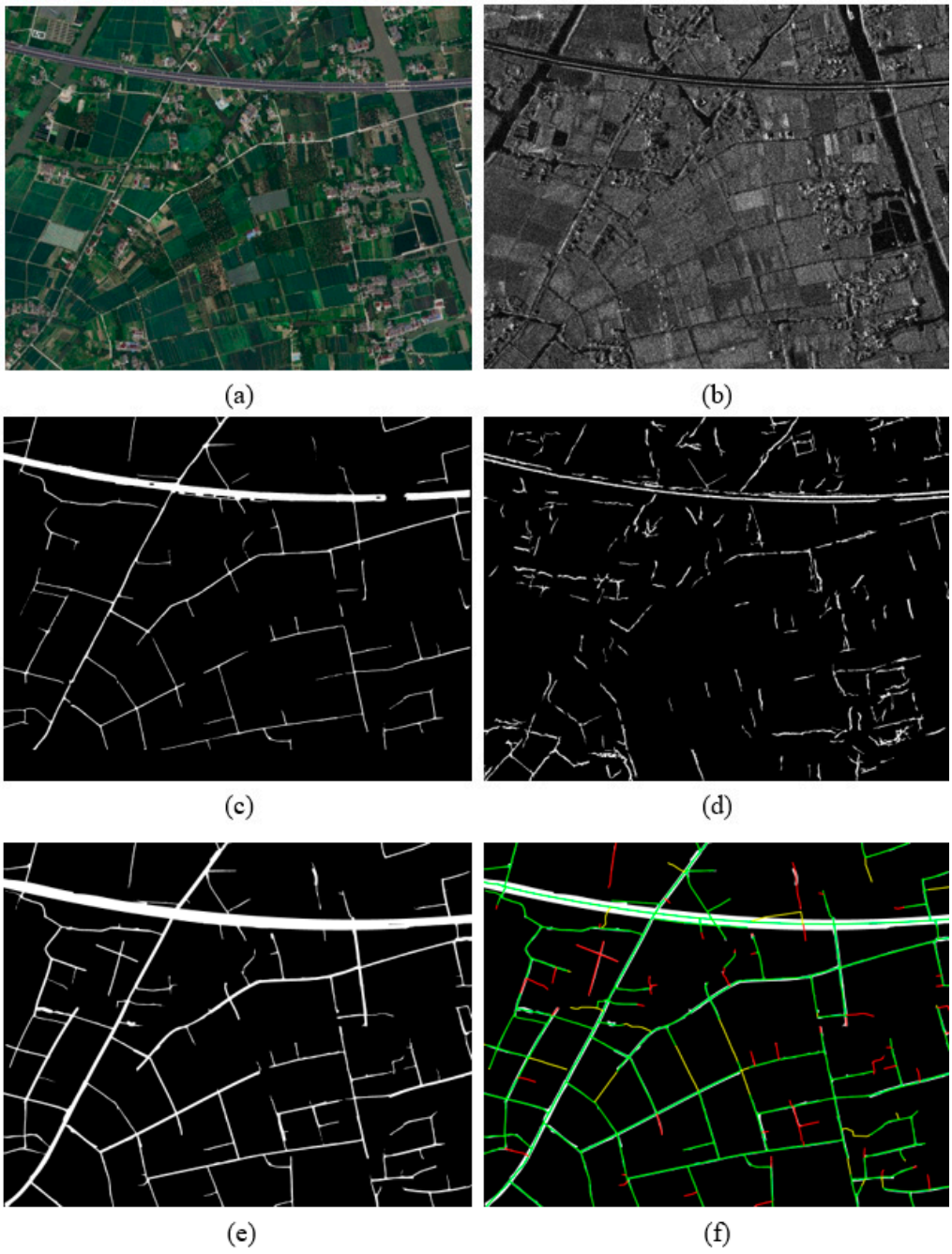



**Figure 24.** Tested data Area 2. Optical and SAR remote sensing images and road extraction results in the Chongzhou area. (**a**) Worldview-2 optical remote sensing image; (**b**) TerraSAR-X remote sensing image; (**c**) road extraction results of optical remote sensing image; (**d**) road extraction results from SAR remote sensing images; (**e**) road fusion extraction results of optical and SAR images; (**f**) ground truth and marking results of road fusion extraction results (green refers to the correctly extracted road, red refers to the incorrectly extracted road, and yellow refers to the omitted real road).

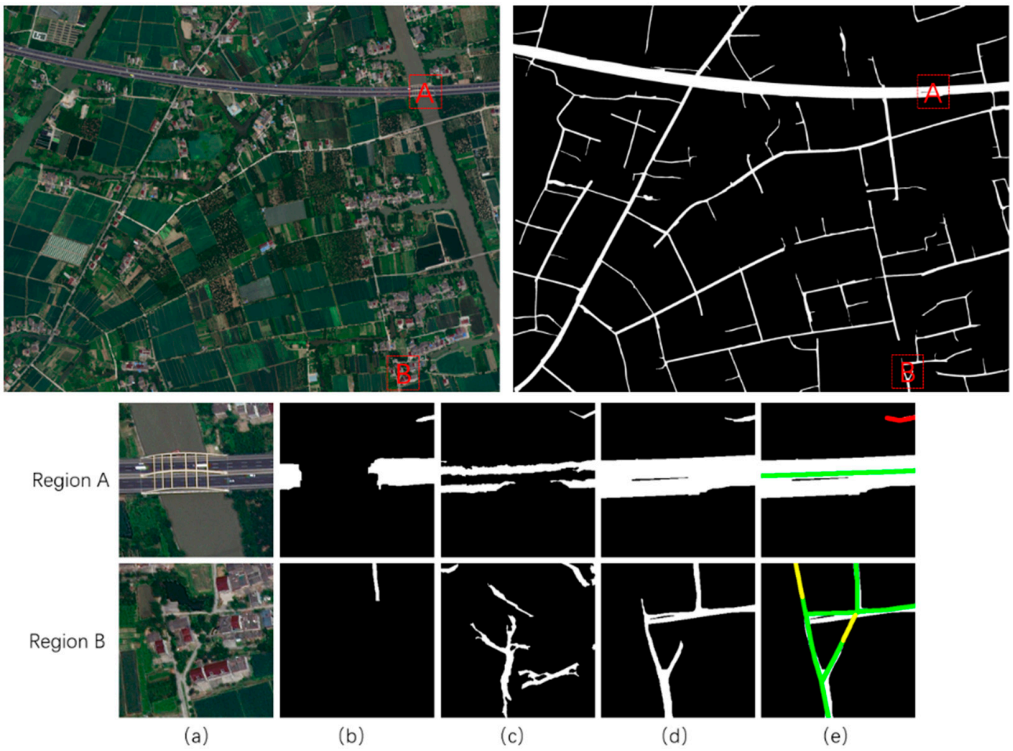



**Figure 25.** Tested data Area 2. Some details in optical and SAR remote sensing images and road extraction results in the Wuzhen area. (**a**) Worldview-4 optical remote sensing image; (**b**) road extraction results of optical remote sensing image; (**c**) road extraction results from SAR remote sensing images; (**d**) road fusion extraction results of optical and SAR images; (**e**) ground truth and marking results of road fusion extraction results (green refers to the correctly extracted road, red refers to the incorrectly extracted road, and yellow refers to the omitted real road).

**Table 6.** The results of two large scale images for the whole area.

| **Tested Data** | **Extraction Results Based on SDG-DenseNet (WorldView Optical Image)** | | | **Extraction Results Based on Decision-Level Fusion Method (SAR and Optical Image)** | | |
|---|---|---|---|---|---|---|
| Metrics | *F*1 | *COR* | *COM* | *F*1 | *COR* | *COM* |
| Area 1 | 0.7376 | 0.8567 | 0.6476 | 0.8528 | 0.9336 | 0.7849 |
| Area 2 | 0.8047 | 0.7923 | 0.8176 | 0.8885 | 0.8680 | 0.9100 |

## 5. Conclusions

In this research, a D-Dense block module was proposed, which combined traditional convolution and dilated convolution based on a dense connection structure. Further, the new semantic segmentation network (SDG-DenseNet) was built with a D-Dense block, and it also adopted the center part of the D-LinkNet for high-resolution satellite imagery road extraction. Since the network also replaces the initial block with the stem block to hold more detailed information, it can be easier to recover the marginal information of the object in the decoder phase. In addition, the introduction of an attention mechanism also improves the ability of the network to obtain global information. Besides, to improve the accuracy of road extraction in large-scale images in practical application, a decision-level fusion method was proposed, which fused the information in optical images and SAR images.

Three sets of satellite images were applied to evaluate the network. The extraction results from the Massachusetts Roads dataset show that the SDG-DenseNet not only has the highest *IoU* and *F*1 score but is also suitable to extract roads in complicated scenes.

Experiments showed that the *IoU* and *F*1 scores of SDG-DenseNet based on D-Dense block and GIRM modules were 3.61% and 2.75% higher, respectively, than the baseline D-LinkNet. The stem block is helpful to develop the accuracy for road extraction. Furthermore, the Chongzhou–Wuzhen dataset, based on three large-scale optical images, was applied to evaluate the models' extraction ability of the low-grade roads. The results show that the SDG-DenseNet performs best in four networks and its *IoU* score is 6.65% higher than that of D-LinkNet. At the same time, its model size is reduced by about 600 MB to D-LinkNet. Further, two pairs of large-scale optical and SAR images were applied to evaluate the decision-level fusion method. The results show that the fusion method performed well in accurately extracting the roads. After decision-level fusion of road binary map from SAR and optical image based on two tested data, the *F*1 is improved by about 8.4–11.5%, *COR* is about 7.4–7.7%, and *COM* is about 9.3–13.7%.

SDG-DenseNet improves d-block as d-blockplus and combines it with an attention mechanism, which not only ensures road completeness in the segmentation task but also greatly improves the correctness of the segmentation results. Therefore, the network maintains a perfect balance between correctness and completeness. In addition, the decision-level fusion method had been proposed to improve the extraction effect on the task of low-grade road extraction, and the presentation quality is better after the decision-level fusion. In future research, the contribution of each part of the network and every hyperparameter in the training phase should be taken into consideration.

**Author Contributions:** Methodology, J.Z.; software, F.X.; validation, Y.S.; formal analysis, B.P.; investigation, J.Z.; resources, Y.L.; data curation, B.P.; writing—original draft preparation, J.Z.; writing—review and editing, S.L.; supervision, Y.L.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Key Projects of Global Change and Response of Ministry of Science and Technology of China under Grant 2020YFA0608203, in part by the Science and Technology Support Project of Sichuan Province under Grant 2021YFS0335, Grant 2020YFG0296, and Grant 2020YFS0338, in part by Fengyun Satellite Application Advance Plan under Grant FY-APP-2021.0304.

**Data Availability Statement:** The authors would like to thank the team of National Climate Center and University of Toronto for the data and experiments.

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

## References

1. 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 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186.
2. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
3. Wei, Y.; Zhang, K.; Ji, S. Simultaneous Road Surface and Centerline Extraction From Large-Scale Remote Sensing Images Using CNN-Based Segmentation and Tracing. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8919–8931. [CrossRef]
4. Yang, F.; Wang, H.; Jin, Z. A fusion network for road detection via spatial propagation and spatial transformation. *Pattern Recognit.* **2020**, *100*, 107141. [CrossRef]
5. Zhou, M.; Sui, H.; Chen, S.; Wang, J.; Chen, X. BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 288–306. [CrossRef]
6. Chen, Z.; Wang, C.; Li, J.; Xie, N.; Han, Y.; Du, J. Reconstruction Bias U-Net for Road Extraction From Optical Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2284–2294. [CrossRef]
7. He, X.; Zemel, R.S.; Carreira-Perpiñán, M.Á. Multiscale conditional random fields for image labeling. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2.
8. Shotton, J.; Winn, J.; Rother, C.; Criminisi, A. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comput. Vis.* **2009**, *81*, 2–23. [CrossRef]
9. Krähenbühl, P.; Koltun, V. Efficient inference in fully connected crfs with gaussian edge potentials. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 109–117.
10. Galleguillos, C.; Belongie, S. Context based object categorization: A critical survey. *Comput. Vis. Image Underst.* **2010**, *114*, 712–722. [CrossRef]

11. Farabet, C.; Couprie, C.; Najman, L.; Lecun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [CrossRef]
12. Eigen, D.; Fergus, R. Predicting depth, surface normal and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2650–2658.
13. Pinheiro PH, O.; Collobert, R. Recurrent convolutional neural networks for scene labeling. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014.
14. Chen, L.C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3640–3649.
15. Lin, G.; Shen, C.; Van Den Hengel, A.; Reid, I. Efficient piecewise training of deep structured models for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3194–3203.
16. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
17. 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; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
18. Lin, G.; Milan, A.; Shen, C.; Reid, I. RefineNet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation. *arXiv* **2016**, arXiv:1611.06612.
19. Pohlen, T.; Hermans, A.; Mathias, M.; Leibe, B. Full-resolution residual networks for semantic segmentation in street scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4151–4160.
20. Peng, C.; Zhang, X.; Yu, G.; Sun, J. Large Kernel Matters—Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
21. Amirul Islam, M.; Rochan, M.; Bruce ND, B.; Wang, Y. Gated feedback refinement network for dense image labeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3751–3759.
22. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
23. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
24. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
25. Hu, C.; Bai, X.; Qi, L.; Chen, P.; Xue, G.; Mei, L. Vehicle color recognition with spatial pyramid deep learning. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2925–2934. [CrossRef]
26. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
27. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
28. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
29. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
30. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 3146–3154.
31. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Computer Vision—ECCV 2018. ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11211.
32. Wan, J.; Xie, Z.; Xu, Y.; Chen, S.; Qiu, Q. DA-RoadNet: A Dual-Attention Network for Road Extraction From High Resolution Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6302–6315. [CrossRef]
33. 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]
34. Sun, Z.; Geng, H.; Lu, Z.; Scherer, R.; Woźniak, M. Review of Road Segmentation for SAR Images. *Remote Sens.* **2021**, *13*, 1011. [CrossRef]
35. Jiang, Y.H.; Pi, Y.J. SAR image road detection based on Hough transform and genetic algorithm. *Radar Sci. Technol.* **2005**, *3*, 156–162.
36. Wei, X.; Lv, X.; Zhang, K. Road Extraction in SAR Images Using Ordinal Regression and Road-Topology Loss. *Remote Sens.* **2021**, *13*, 2080. [CrossRef]

37. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 2–4 February 2017.
38. Rahman, M.A.; Wang, Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 12–14 December 2016; pp. 234–244.
39. Jégou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
40. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Orr, G.B.; Muller, K.R. *Efficient Backprop in Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 9–48.
41. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
42. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]
43. Xiao, F.; Chen, Y.; Tong, L.; He, L.; Tan, L.; Wu, B. Road detection in high-resolution SAR images using Duda and path operators. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1266–1269. [CrossRef]
44. Mnih, V.; Hinton, G.E. Learning to Detect Roads in High-Resolution Aerial Images. In Proceedings of the Computer Vision—ECCV 2010—11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Proceedings, Part VI; Springer: Berlin/Heidelberg, Germany, 2010.
45. Sun, T.; Chen, Z.; Yang, W.; Wang, Y. Stacked u-nets with multi-output for road extraction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 202–206.
46. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]
47. Zhang, L.; Lan, M.; Zhang, J.; Tao, D. Stagewise Unsupervised Domain Adaptation with Adversarial Self-Training for Road Segmentation of Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5609413. [CrossRef]
48. Zhang, Z.; Wang, Y. JointNet: A common neural network for road and building extraction. *Remote Sens.* **2019**, *11*, 696. [CrossRef]
49. Dey, M.S.; Chaudhuri, U.; Banerjee, B.; Bhattacharya, A. Dual-Path Morph-UNet for Road and Building Segmentation From Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3004005. [CrossRef]