remote sensing

MDPI

Article

WSGAN: An Improved Generative Adversarial Network for Remote Sensing Image Road Network Extraction by Weakly Supervised Processing

Anna Hu [1], Siqiong Chen [2], Liang Wu [2], Zhong Xie [2], Qinjun Qiu [1] and Yongyang Xu [2,*]

[1] National Engineering Research Center of Geographic Information System, Wuhan 430074, China; huanna@cug.edu.cn (A.H.); qiuqinjun@cug.edu.cn (Q.Q.)

[2] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China; chensiqiong@cug.edu.cn (S.C.); wuliang@cug.edu.cn (L.W.); xiezhong@cug.edu.cn (Z.X.)

* Correspondence: yongyangxu@cug.edu.cn

check for updates

Citation: Hu, A.; Chen, S.; Wu, L.; Xie, Z.; Qiu, Q.; Xu, Y. WSGAN: An Improved Generative Adversarial Network for Remote Sensing Image Road Network Extraction by Weakly Supervised Processing. *Remote Sens.* **2021**, *13*, 2506. https://doi.org/10.3390/rs13132506

Academic Editor: Saeid Homayouni

Received: 11 May 2021
Accepted: 23 June 2021
Published: 26 June 2021

Abstract: Road networks play an important role in navigation and city planning. However, current methods mainly adopt the supervised strategy that needs paired remote sensing images and segmentation images. These data requirements are difficult to achieve. The pair segmentation images are not easy to prepare. Thus, to alleviate the burden of acquiring large quantities of training images, this study designed an improved generative adversarial network to extract road networks through a weakly supervised process named WSGAN. The proposed method is divided into two steps: generating the mapping image and post-processing the binary image. During the generation of the mapping image, unlike other road extraction methods, this method overcomes the limitations of manually annotated segmentation images and uses mapping images that can be easily obtained from public data sets. The residual network block and Wasserstein generative adversarial network with gradient penalty loss were used in the mapping network to improve the retention of high-frequency information. In the binary image post-processing, this study used the dilation and erosion method to remove salt-and-pepper noise and obtain more accurate results. By comparing the generated road network results, the Intersection over Union scores reached 0.84, the detection accuracy of this method reached 97.83%, the precision reached 92.00%, and the recall rate reached 91.67%. The experiments used a public dataset from Google Earth screenshots. Benefiting from the powerful prediction ability of GAN, the experiments show that the proposed method performs well at extracting road networks from remote sensing images, even if the roads are covered by the shadows of buildings or trees.

Keywords: weakly supervised; road extraction; remote sensing image; generative adversarial networks

1. Introduction

With the rapid development of observation and sensor technology, there has been a gradual improvement in the spatial resolution of remote sensing images, which provide substantial data sources for road extraction research [1,2]. Road networks extracted from remote sensing images have been applied in several fields, such as geological surveys [3–5], disease monitoring [6–8], and analysis research [9–11]. As remote sensing images are timely updated, the extracted road networks can be beneficial for road planning [12,13], construction, graphic similarity measure [14], and maintenance [15,16], and they can help update road network information. As the satellites are overhead and view sensors and the remote sensing images are large, road extracting methods must overcome some problems, such as the efficiency problem, shelter occlusion problem by trees (or other objects), and the data preparation problem during the training process.

(1) Efficiency problem. As in other remote sensing image object extraction research, efficiency is very significant to the large quantity of remote sensing image resources. Initially,

the experts used computer vision methods to extract specific road features, such as straight segments and bends. Those studies [17,18] used spectral information and topological relations, such as the line feature detection algorithm based on random transformation [19] and aperiodic directional structure measurement (ADSM) to extract road networks [20] and distinguish between road networks and other ground objects. These methods can accurately obtain the single road features, such as a straight road segment and a bending road segment, when their parameters are manually set to distinguish every ground object feature. These methods can accurately extract road networks only in some special remote sensing scenes. Almost all remote sensing images have various complex scenes; the single-feature extract structure cannot completely and effectively express the characteristics of the road network. Moreover, manually setting the parameter is time-consuming and subjective.

(2) Shelter occlusion problem. As outlined above, the single-feature extraction method is not enough for remote sensing images for road extraction; therefore, this study needs an appropriate multi-feature extraction technique. Some researchers have used machine learning road extraction methods, which use classification or segmentation methods to extract road networks and distinguish them from other objects [21–23]. In traditional machine learning methods, the support vector machine (SVM), K-means, and other methods are used to capture road network features. For example, Song et al. [24] proposed an SVM trained using the spectral information in a remote sensing image to extract roads. The K-means [25] iteration was designed for the segmentation of images, where the class of a road can be recognised, and the recognition result is filtered using a mathematical morphology method. A new conditional random fields (CRF) method [26] was designed to detect road marks, in which prior information is presented by high-order pixel patches, which are connected by straight line super-pixels. In deep learning methods, the powerful feature extraction neural network method uses a large number of parameters to enhance the ability of the multi-feature extraction. Xu et al. [27] designed a semantic segmentation method based on an end-to-end neural network, which is an improved U-net [28] framework. It employs an attention patch for semantic segmentation modules to learn global and local features. Zheng et al. [29] used a CRF-RNN neural network to extract road networks from remote sensing images, which transfers a loss to optimise the model via the deep learning method's widely used error transform method, backpropagation (BP) algorithm, and obtained satisfactory results. Compared with computer vision road extraction methods, traditional machine learning road extraction methods have more powerful feature-extraction abilities. However, those methods are still not enough for the remote sensing image to extract road network from complex ground object scene. The traditional machine learning road extraction methods need structural features (straight, circumplex and so on) by artificial methods, such as the adjustment loss function; it is not suitable for the remote sensing image, which shelters by trees and buildings. Machine learning methods cannot predict road networks from occluded remote sensing images [30]. Accordingly, a novel image feature prediction method should be considered for remote sensing image extraction.

(3) Data preparation problem. As discussed above, the machine learning methods need more powerful feature prediction capabilities to recover occluded sections of road networks. Fortunately, the generative adversarial network (GAN) [31], which uses a new training strategy that can avoid artificial errors from the loss function, was proposed. This framework is adequate for learning deep features from training data and has a certain predictive ability. For instance, the paper [32] considers the problem of cloud cover during road extract and designs an improved GAN network that combines an edge prediction framework with a colour filling part. There are also studies that used conditional generative adversarial network (CGAN) to construct semantic segmentation framework for high-resolution remote sensing images [33]. This method proposed a segmentation method that combined long short-term memory (LSTM). This method is named CGAN-LSTM and it uses ground truth as the constraint condition to enhance high-level spatial

information for remote sensing images. The results show that GAN has a powerful ability with shape prediction for ground objects. Benefiting from the adversarial training strategy of GAN, these segmentation methods have outstanding performance in terms of precision and further reduce the influence of occlusion from buildings and trees. However, these methods almost used supervised training data to fit the parameters, which needs labeled segmentation images to constrain the training process [34]. The road segmentation labels need artificial marking, and remote sensing images are complex. Therefore, these methods take a lot of time and resources. A novel data preparation method should be considered for remote sensing images for road extraction. Due to the adversarial strategy of GAN, it is difficult to balance the training process between the generator and the discriminator. The standard GAN loss is not enough for the GAN training process.

Accordingly, to achieve road extraction more precisely, two issues should be solved: (1): the problem of road extraction labels needing artificial marking for building the training dataset. (2): balance the influences in loss optimisation during the training process. Thus, following the research line of relating works, this study introduces some outstanding methods to promote the performance of the road network extraction algorithm. The main contributions of this study are the following:

(1) This study proposes a weakly supervised method based on GAN to extract road networks from remote sensing images. The weakly supervised method significantly improves the dataset preparation process compared with the current supervised method [35], which uses remote sensing and mapping images to train the model, rather than pairs of remote sensing and binary images.

(2) Owing to the outstanding prediction ability of GAN, the proposed method can further overcome masking problems due to buildings and trees, allowing the detection of a clear and straight road network. Furthermore, Wasserstein GAN with gradient penalty loss (WGAN-GP) [36] is used to strengthen the efficiency of GAN, which avoids model collapse.

(3) Residual network (Res-net) block [37] is used to enhance the feature extraction ability from the complex remote sensing scene, which combines shallow information with high-frequently information and does not miss the texture information.

The remainder of this paper is organised as follows. Section 2 presents the methodology used for this study, including the GAN mapping framework, the loss in the GAN framework, and the post-processing operations. Section 3 presents the experimental results. Section 4 analyses the results of the experimental and focuses on group discussions. Section 5 provides our conclusions.

2. Road Extraction from Remote Sensing Imagery

Weakly supervised road extraction has a large focus, where relatively few studies have been devoted to remote sensing image processing. Due to limitations in training datasets, which need pairs of segmentation images to be obtained, supervised road network extraction methods still have some problems. Weakly supervised methods can alleviate onerous demands of preparing the training data and reduce the requirements for the generation of road networks for which generating very accurate segmentation results is required. Thus, this study proposes a weakly supervised method based on GAN. This extraction method is completed in three steps. In a mapping network, the remote sensing image can be converted into a mapping image. In the complex remote sensing scene, to restore the road network of shelter by the tree, the information in no tree's road is very important. Thus, to extract more features from the global information (other place's information in original feature map), the mapping introduces numerous Res-net blocks. Second, in the model optimisation process, the instability loss function is replaced by WGAN-GP. Finally, in the post-processing process, the mapping image is converted into a binary image by a threshold value. To obtain more precise results, this study used a morphological method, known as the dilation and erosion method (DE) [38], to remove salt-and-pepper noise [39]. The structure of the road extraction network is shown in Figure 1.

Especially, to clearly display the structure of mapping image generator, discriminator, and Res-net, multiple colours are used in Figure 2, and Figure 4, which can clearly distinguish each feature map in feature map block. In Figure 3, the different colour means a different feature map block that is calculated by the convolution process.

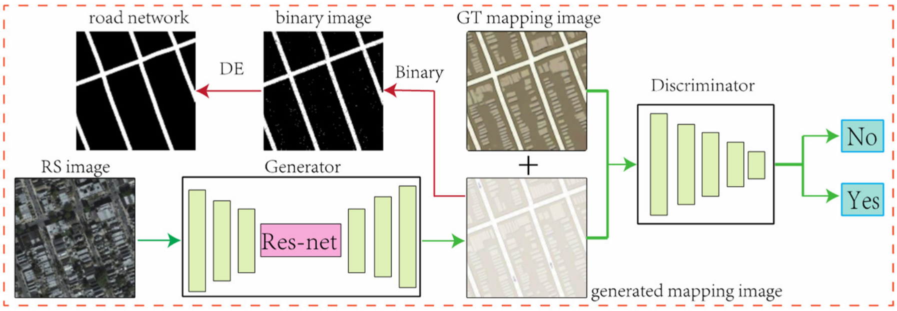

Figure 1. Road extraction network structure.

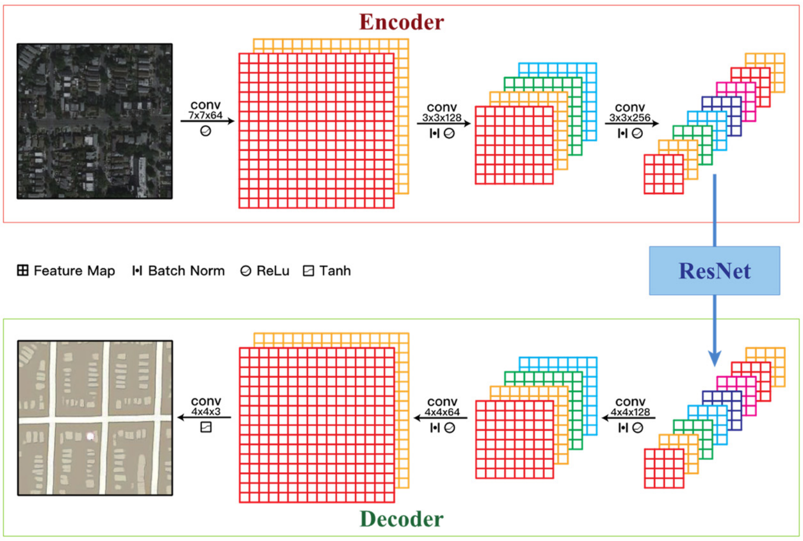

Figure 2. Structure of the mapping network.

2.1. WSGAN Model Structure

This study designed a road network extraction method, referred to as WSGAN, which is based on GAN and transforms the remote sensing image into a mapping image using a weakly supervised method, followed by the extraction of the road networks from the generated mapping image. The advantage of the mapping network can be summarised as follows: the GAN framework can learn a loss function that self-adapts to the training dataset and avoids errors related to unsuitable loss functions. This strategy alleviates the problem of artificial design loss function error present in most deep neural networks; it is powerful enough in the image shadow area. Encouraged by the performance of the GAN for semantic segmentation, the WSGAN was designed in two parts. The first part is the generator, *G*. In the WSGAN framework, *G* is used to generate the mapping image and mislead the discriminator, *D*, which is the second part designed to discriminate the original mapping image from the synthesis image. Both parts are trained by a confrontation strategy. In structure G, the outstanding data normalisation strategy, BN, is used in each convolution processing. Moreover, the ReLU and Tanh are used in G and D during the BN process.

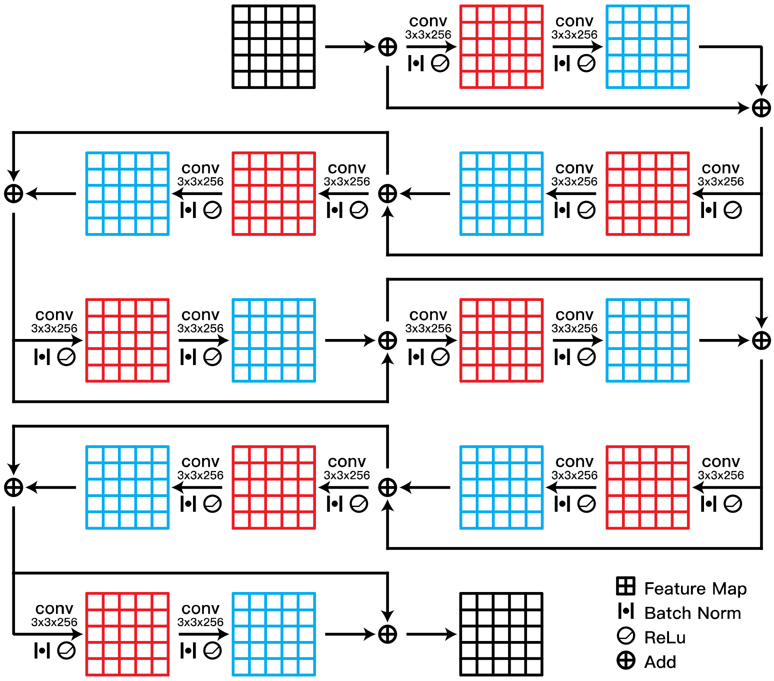

Figure 3. The Res-net block structure between encoding and decoding.

In many studies, the generator framework is composed of traditional convolution neural networks, such as end-to-end networks. However, these end-to-end networks are not sufficient for the extraction of features from road networks, which cannot restore the global information for the deep layer. Thus, the skip connections in the Res-net block can greatly restore the information in the shallow network. Res-net is efficient at extracting features from images, such that, in the generator, G, this study used numerous Res-net blocks to extract the road network features. Figure 2 shows the details of the mapping image generation network.

The Res-net blocks employ a learning strategy termed residual learning. In previous studies, the deep learning neural network uses a non-linear method to transform input information, while Res-net proposes a new connection method termed skip connections. Compared with traditional convolution neural networks, Res-net blocks allow original information to be connected with the previous layer. Moreover, traditional convolution and fully connected frameworks easily lose information during transmission. Vanishing and exploding gradients in convolution networks hamper the training of the deep network. The parameters can be reduced in this network, and the learning target can be more easily achieved. For remote sensing images, the residual learning method further restores high-frequency feature information for road networks and ensures that the G network can recognise more prominent road objects. Moreover, the skip-connected strategy can restore the global information (position information and other shallow features) to a great extent. The structure of Res-net accelerates the training process and improves the accuracy of the model. Figure 3 shows the details of the Res-net.

Despite their efficacy extract results, traditional neural networks that use a self-design optimisation loss function to fit models suffer from a major drawback, i.e., the large error used by the artificial loss function, which needs the road network extract experience and through that subjective information to complect the research. To solve this problem, this study used GAN to learn the loss function using a discriminant network for each training dataset. The discriminator of GAN is a classification network that distinguishes the original mapping images from the synthesis images. Thus, through optimising the loss function, the discriminator can accurately find the difference between the original and the synthesis mapping image. Compared with the artificial design optimisation function, the discriminator can automatically learn the optimised method through this training data. Thus, the GAN can avoid artificial errors due to the loss function, which does not need to design feature loss function by subjective experience and cause the inaccuracy model optimisation.

To improve the ability of the discriminator, this study used a training strategy known as Patch GAN [40], which used average value to replace the original sigmoid results. PatchGAN removes the sigmoid strategy and mapping the feature map as an N×N matrix. The matrix will serve as the score to evaluate the authenticity of generated mapping image. This strategy is considered the receptive field as a convolution network, and each value of the matrix represents the score of each part in generated mapping image. Compared with the original discriminator, PatchGAN greatly focuses on more field in generated mapping image. This strategy considers the effects of different parts against the image rather than discriminating the image as a whole. A recent study, employing Pixel2Pixel [41], showed the effectiveness of this strategy by comparing the results of patches of different sizes. For an input image of 256 × 256 pixels, when the patch size was 70 × 70 pixels, the results were similar to the input image. The Patch GAN can reduce the parameter size to improve the operational efficiency, discriminate the details in every patch, and enhance the ability of the generator. Owing to the independence of each patch, *D* will generate different results, where the final result is the averaged value of several intermediate results. Figure 4 shows the structure of the discriminator.

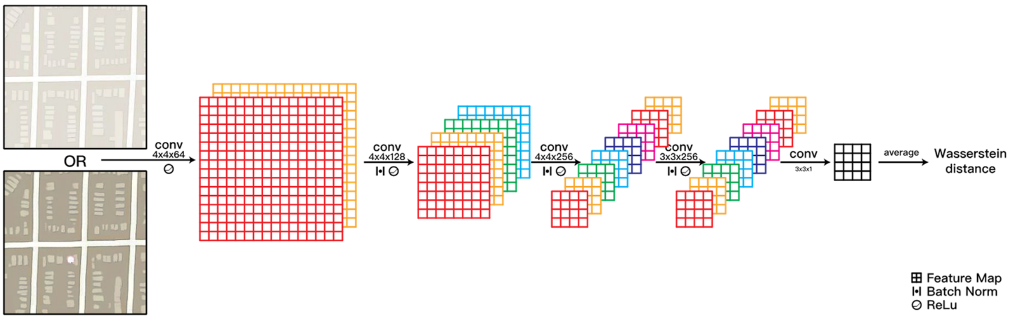

Figure 4. The structure of the discriminator *D*.

2.2. *Mapping Model WGAN-GP Optimisation Processing*

The loss function was used to control the training process of the generator and discriminator. The GAN training strategy is a promising approach for the learning of image process models. This method first uses a combat strategy for training a neural network, which aims to reduce the distance between the synthesis and target images. Despite its enduring success, the structure of GAN has several problems. First, due to training by an adversarial strategy, using the GAN loss function to balance the level of training of the generator, *G*, and discriminator, *D*, is difficult. In previous studies, which were based on the standard GAN method, the training step was carefully designed through the experience gained until the convergence of the model. Second, during training, the generator always suffers from the model collapse phenomenon, where similar results are produced by the generator. From this, during the training process, the experiments based on standard GAN loss function would cause disordered results and broken outputs. Finally, confirming whether the training process converged is challenging. Thus, the standard GAN loss function cannot optimise the parameters for road extraction. In the original GAN framework, the adversarial loss function converges by minimising the KL distance (Kullback–Leibler divergence) [42] and maximising the JS distance (Jensen–Shannon divergence) [43], which causes the generator to produce sample images with high diversity, but low quality.

There are a lot of studies that have designed workarounds to solve this problem, such as WGAN-GP and WGAN [44]. For WGAN, the Wasserstein distance has been used to calculate a new loss function, known as WGAN loss. However, the weight of the generator is required to clip results in extreme parameter values. This problem causes poor results and gradual data overload during training. Besides, gradient punishment is combined with weight clipping in WGAN-GP. This loss function further stabilises the performance of WGAN. Therefore, the WGAN-GP loss function is a weight control method in WGAN that restrains the weight of the discriminator via a gradient loss function. WGAN-GP is an effective strategy for controlling the training progress through the weight control loss in the discriminator function. For this, WGAN-GP loss replaced the original standard GAN loss and was used to optimise the discriminator. The optimisation algorithm of the mapping model is as follows:

$$\underset{G}{min}\underset{D}{max}V(G,D) = E_{y\sim P_{data}(y)}(D(y)) - E_{x\sim P_{data}(x)}(D(G(x))) + \lambda E_{x\sim P_p}[\| \nabla_x D(x) \|_p - 1]^2 \tag{1}$$

where $p = \varepsilon x_r + (1-\varepsilon)x_g$, x_r is the distribution of the real mapping image, x_g is the distribution of the generated mapping image, and ε is a random value between 0 and 1.

In the training process of *D*, the parameter of *G* is fixed, and the parameters of *D* will be optimised by *D* loss. The purpose of *D* is distinguished generated mapping image with GT mapping image; thus the *D* loss would make the bigger value in the score of *D(y)* and the smaller in the score of D(G(x)). Especially, to complect the strategy of WGAN-GP, the gradient penalty is also added in *D* loss. The *D* loss function is as follows:

$$\underset{D}{max}V(G,D) = E_{y\sim P_{data}(y)}(D(y)) - E_{x\sim P_{data}(x)}(D(G(x))) + \lambda E_{x\sim P_p}[\| \nabla_x D(x) \|_p - 1]^2 \tag{2}$$

In the training process of *G*, the is fixed, and the parameters of *G* will be optimised by *G* loss. The purpose of *G* is made *D* cannot distinguish generated mapping image with GT mapping image; thus the G loss would make the bigger value in the score of *D*(*G*(*x*)) and the smaller in the score of *D*(*y*). Especially, because the fixe of the parameter of *D*, $E_{y\sim P_{data}(y)}(D(y))$ is fixed. Thus, the *G* loss is as follows:

$$\underset{G}{min}V(G,D) = -E_{x\sim P_{data}(x)}(D(G(x))) \tag{3}$$

Moreover, to improve the similarity of the generated mapping images to the original mapping images in the pixel feature, this study used the L_1 loss to optimise the generator. The complete mapping network loss function is as follows:

$$L = arg\underset{G}{min}\underset{D}{max}V(G,D) + \lambda L_{L_1}(y, y^{\sim}) \tag{4}$$

where λ is the weight of the L_1 loss, *y* is the original mapping image, and $y^{\sim}$ is the generated mapping image.

2.3. Binary Image DE Method Post-Processing

Spectra are the specific attributes of the ground object in a remote sensing image, where each type of ground object corresponds to a spectral curve. The difference in the reflected or emitted battery radiation energy in different wavebands of ground objects is due to the formation of a colour difference in the RGB image. Thus, according to the spectral characteristics of remote sensing images, road networks can be effectively extracted. In this study, remote sensing images were transformed into mapping images, which had a uniform distribution of pixel values, where the road network area had a value of 255 and other objects had a value of 0. For the mapping image, binarisation processing was used to extract the road network. A dynamic adjustment of the threshold value, selected through the binary results, was used to segment the mapping image.

Image binarisation transforms the mapping image into a binary image, whose pixel values are 0 or 255. All pixels whose values are equal to or above the threshold value were judged as belonging to the road network, with a grayscale value of 255; otherwise, pixels with values below the threshold were excluded from the road network, with their values set to 0. These areas represented the background or other objects. Thus, the mapping image can be presented as a binary image, where the road network is notable. In other words, the values of the mapping images were transformed to extreme values by an appropriate threshold value while the binary image retained the road network features in the mapping image. Binary images play a critical role in the field of digital image processing, significantly reducing the amount of data for the mapping image and highlighting the outline of the road network. At the same time, the image must be binarised to perform binary image processing or analysis.

Despite their safety and efficacy, binary images produced by a threshold value always suffer from salt-and-pepper noise, also known as impulse noise, which is a random occurrence of white or black dots in otherwise white or dark areas, as well as a light and dark point noise, which is checkered with black and white. In most instances, the image sensor, transmission channel, or decoding process produces salt-and-pepper noise. Black noise spots are known figuratively as pepper noise, while white noise spots are known as salt noise. In general, these two types of noise always appear simultaneously. Salt noise may be caused by a sudden and strong interference in the image signal during conversion from analogue to digital or during bit transmission, where a failed sensor results in a minimum pixel value (black) and a saturated sensor results in a maximum pixel value (white).

At present, morphology methods are used to remove salt-and-pepper noise. To obtain better experimental results, this study compared several morphological removal methods that can process noise under complex circumstances. The DE morphology method is efficient at removing the noise from a binary image and uses structural units to measure and extract corresponding patches to achieve image analysis and recognition. This method can effectively remove noise and retain the original features of remote sensing images.

The DE method is known as a morphological operation, which is typically used to remove noise from binary images and is similar to the contour detection method. The dilation process adds pixels to the boundaries of perceptual objects and expands the bright white areas in the enlarged image. This process allows the boundary point to shrink into the object's interior and removes small noise points, such as salt noise. In contrast, erosion removes the noise and reduces the object size along object boundaries. This process combines the background point, which is in contact with the object, to fill the mask. The input for the noise removal method is the binary remote sensing image, while the output is the binary road network image, which is clear and straight. Figure 5 shows the results of this method.

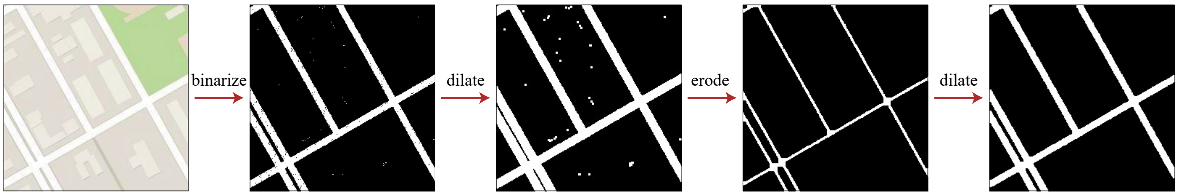

Figure 5. Schematic of post-processing.

3. Results

3.1. Dataset

The training dataset of the mapping network was composed of a remote sensing image and the corresponding mapping image, which were collected by a Pixel2Pixel model. The datasets were screenshots from Google Map, and the images were sampled from in and around New York City, which has obvious shelter from trees and can make the road extraction more difficult. Especially, the straight road network can make WSGAN focus on solving the problem in restoring the no-cut-off road network. In GAN mapping, remote sensing images were transformed into mapping images. Thus, the dataset from Pixel2Pixel was suitable for our experiments. This characteristic is beneficial to confirm our method's performance, erasing the influence of any masking. Moreover, in this dataset, the colours of water and forest areas differ from those of buildings. Owing to the purpose of this study, the road network was the most important factor during the training process. Thus, water and forest, covering large areas, were deleted. After this pre-processing, the training dataset was composed of 5336 remote sensing and 5336 mapping images. Additionally, the test data was composed of 106 pair remote sensing images and mapping images. Moreover, the ground truth of the segmentation image was obtained by an appropriate threshold. To meet the input requirements for the mapping network, the training images were cut to 256 × 256 pixels from 512 × 512 by clipping via a sliding window. During training, the input for the discriminator was generated through the mapping and the original mapping image. Due to the discriminator structure in Patch GAN, the connected image was cropped with patches of 128 × 128 pixels. Especially, in every figure, the results images are selected from huge test results, and the evaluation scores in each table are calculated by all test images. Figure 6 shows the training data for the mapping image and the binary processing results.

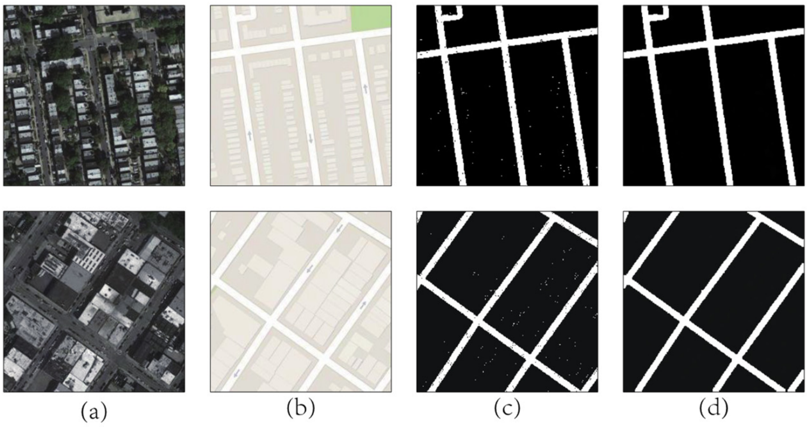

Figure 6. Example of the model's training dataset: (**a**) remote sensing image, (**b**) mapping image, (**c**) binary image with the salt-and-pepper noise, and (**d**) extracted result.

3.2. Experimental Setup and Results

In this model, the road extraction process was performed in three steps. First, the remote sensing images were transformed into mapping images by the WSGAN model. The generated mapping images were similar to the original mapping data from Google Earth, which has a clear outline. Second, the generated mapping images were transformed into binary images, which are black-and-white images, where white indicates the road network and black indicates other objects. Finally, the binary image always suffered from salt-and-pepper noise, which was removed by the DE morphology method. Figure 5 shows the post-processing results. Moreover, the size of all samples was 256 × 256 pixels, rendering the data flow simple. To reduce the number of parameters and improve the operating efficiency of the mapping network, the discriminator, D, used the strategy of the Patch GAN, which cropped the input for the discriminator and averaged the output for the final result. Here, the patch size at the end of the discriminator was 128 × 128 pixels, representing a quarter of the original image. During training, the mapping network was optimised by the root mean square prop (RMSprop) method, which is suitable for combat training. The learning rate was set to 0.0002, the batch size was 1, and we performed around 200 epochs on each dataset to ensure convergence. We alternated updates to the generator and discriminator networks, where the adversarial loss weight was 1, the gradient penalty weight was 10, and the L_1 loss weight was 10. We trained our model on a Linux OS with four NVIDIA Titan XP GPUs and 12 GB of RAM.

The main challenge faced by experiments using the road extraction method is the shadows of trees and buildings. This problem influences the accuracy of the road extraction and generates an incomplete road network that affects its actual usability. The structure of the used GAN, however, has a strong predictive ability for incomplete objects. Furthermore, to structure a more general model, this method used a weakly supervised method to train the model, transforming the remote sensing images to the mapping images without the binary image. The mapping image was available on Google Earth, while the binary image required artificial creation. Based on this, this study used our proposed method to extract the road networks, which can solve the problem of shadows from trees and buildings. As shown in Figure 7, the results of the mapping network had complete edges.

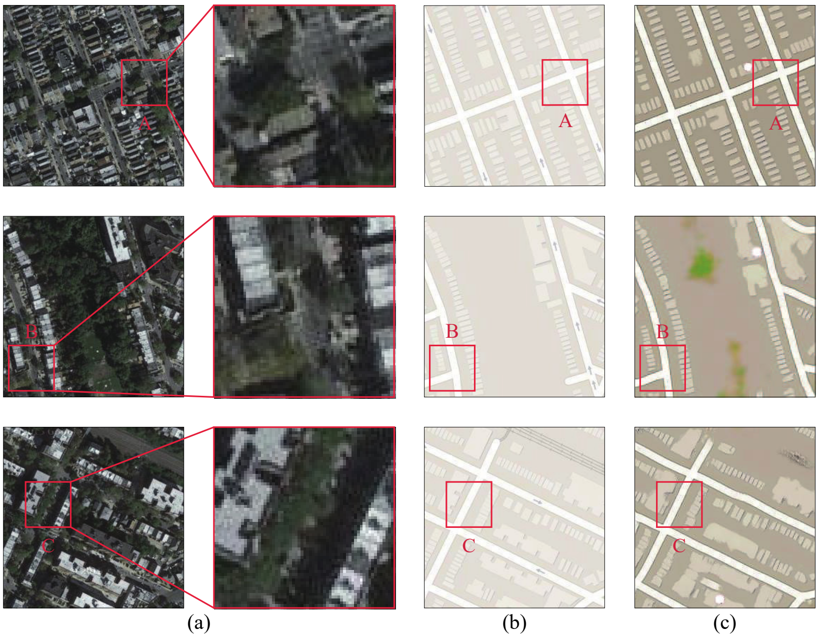

Figure 7. Results of the mapping network: (**a**) remote sensing image and the shelter area by tree; (**b**) GT and (**c**) synthesis images; (A), (B), and (C) are the patches in the same areas.

As shown in Figure 7, in each patch, the remote sensing image is obscured by a large ground object, such as buildings or trees. In this study, three images, extensively covered by ground objects, were selected in each patch. Comparing the generated mapping images with the original mapping images, the results for the road network were similar to the ground truth, despite the fact that the road network was obscured by many trees or buildings. The GAN framework is generally accepted as yielding outstanding performance for the image transform task [45–47]. This is consistent with our results shown in Figure 7. The mapping network was efficient at the extraction of the road network from the remote sensing images.

To select the best threshold, this study implemented the following steps. First, the real mapping image had a defined threshold in Google Earth by which the mapping image was transformed into the binary image. This binary image represented the ground truth (GT) to calculate the threshold value. Second, the generated mapping image was binarised by numerous thresholds, where each of these binary images was denoised by the morphological method. Finally, to evaluate whether the threshold value was optimal, this study calculated evaluation indices between the GT and the generated binary image, such as the IoU (intersection over union), P (precision), R (recall), ACC (accuracy), and F1 (harmonic mean). Based on Figure 8, the best threshold was between 210 and 220, where 213 was the best value because the IoU (0.85), ACC (0.98), p (0.90), R (0.90), and F1 (0.92) reached their maximum values. Moreover, the BEP (Break-Even Point), in which the precision and recall are the same, indicated the optimal value. Based on this threshold, the binarisation was performed.

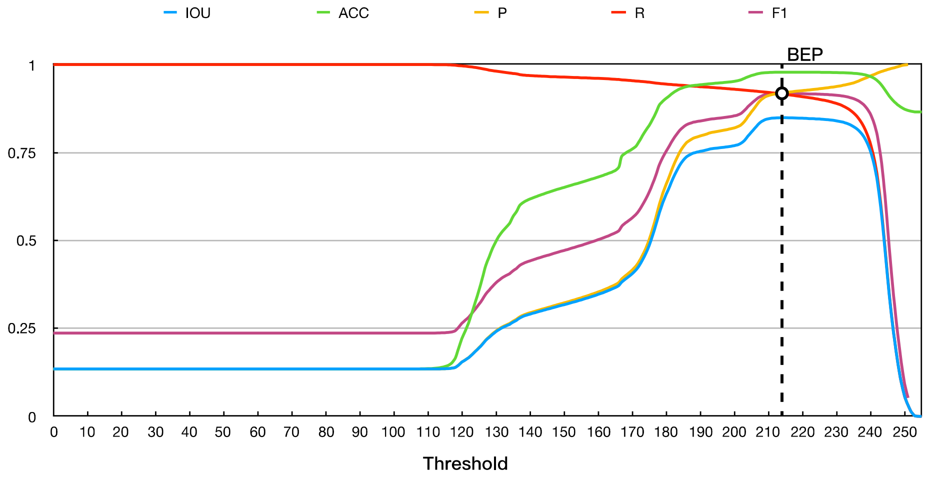

Figure 8. Results of the evaluation indices at each threshold.

To show the performance of this method, this study selected three remote sensing images from several test results to display the road extract results, which exist visually obvious masked in many road network areas by trees and buildings. As shown in Figure 9, to binarise the mapping image, the mapping network transformed the remote sensing image with a stronger colour contrast for each object. For this, the colour contrast in the remote sensing image can be easily segmented.

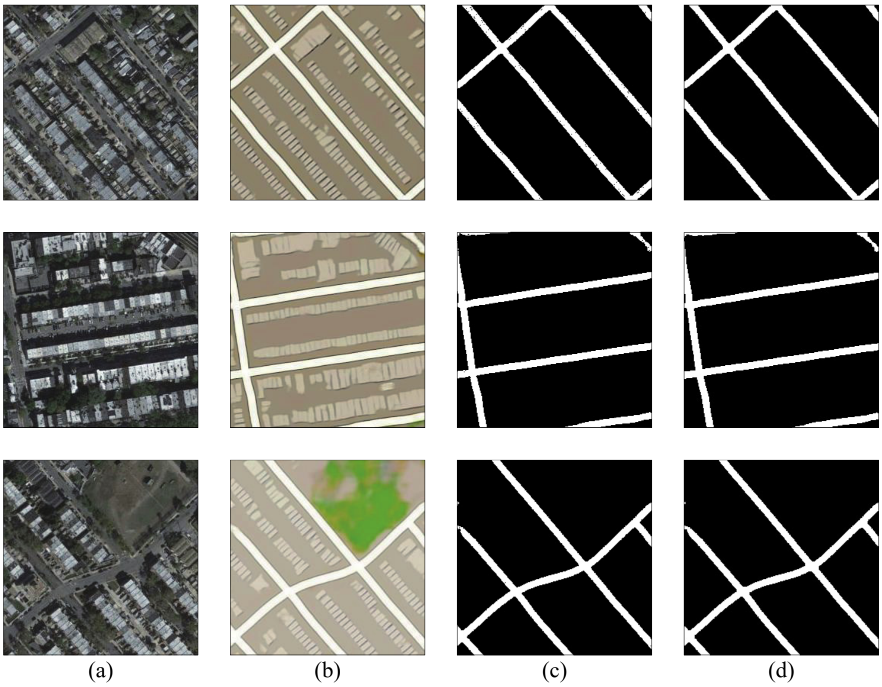

Figure 9. Results for the road network: (**a**) remote sensing images, (**b**) synthesis images, (**c**) binary images, and (**d**) denoised images.

Based on the results, masking by trees and buildings can be forecast by the WSGAN, where the mapping image presented a sharp road network. To highlight the road network, the mapping model generated clear colour mapping images, in which the roads were white and other objects were black. All of the processing is shown in Figure 9. Our results indicate that our method performed well at separating roads from other ground objects. Moreover, more notable road network colors distinguished the network from other objects, yielding higher accuracy for the results. This study enhanced the chromatic with ground truth and produced less salt-and-pepper noise.

4. Discussion

Due to the limitation in the recent road network extraction method, this study used a weakly supervised method, which had three steps: mapping, binarisation, and removal. Furthermore, to obtain more robust results, in the basic GAN model, this study made the following improvements to the mapping network. First, to obtain a complete road network and enhance the predictive ability of the generator, Res-net was used for the generator to improve the feature extraction ability. Second, during the training process, the structure of the GAN has problems converging, and the generated results become uncontrollable, a phenomenon termed model collapse. The WGAN-GP was effective at avoiding model collapse as opposed to the GAN loss during the training process. In contrast to the standard GAN, the WGAN-GP produced a weight control strategy and used the gradient control method to clip the value of the weight. For the contributions of this road network extraction method, the salt-and-pepper noise removal methods have been discussed, and the efficiency of the innovation factor was compared. Moreover, the threshold advantage of this method is fully analysed.

In Figure 10, to accurately extract the outline of the road network, this study compared the DE method to previous studies on noise removal. Additionally, the details about the denoise results are magnified at the bottom of the results. The results of the DE were the most accurate and most similar to the ground truth. The results of the other methods showed excessively smooth edges, such that the boundary of the road network appears unrealistic, which reduces the utility of the road network extraction model and reduces the efficiency of the binary image. As shown in Figure 10, the salt-and-pepper noise removal method showed strong robustness, which performed well on the testing dataset. The edge of the road network binary image is evident. The pepper noise in the road network and the salt noise in other objects were removed. The class of the road network with other ground objects became more notable. Therefore, our model can remove the salt-and-pepper noise from binary remote sensing images.

Moreover, to prove the effectiveness of our method, the innovation factor was compared with the binary result. The mapping images were binarised by the DE method, and the mapping methods were trained by the same training dataset, whose results are shown in Figure 11.

As shown in Figure 11, WGAN-GP had an outstanding performance and avoided the model collapse phenomenon. Moreover, Res-net was efficient at feature extraction, such that there was improved detection of the road network. Compared with the other methods, the WGAN-GP loss had a powerful ability with respect to image generation. The weight clip method produced an appropriate network weight rather than clipping the weights by extreme values. Moreover, in masked areas, our method recovered a sharper mapping image. The road network had straight edges in the WGAN-GP results, where the colour of the road network was distinguishable from those of other objects.

In most instances, the road network had similar spectral information with other ground objects. Therefore, although this study generated the mapping image from a remote sensing image, the road networks pose challenges when extracting them from mapping images. Thus, to extract a precise road network, the mapping image generated model must have a sensitive spectral information perception ability. To evaluate the spectral information perception ability, Figure 12 compares our results with the improved method's results (WGAN loss and Res-net) for each threshold, which had identical scores before the pixel value was set to 100.

As shown in Figure 12, this study compares the scores of evaluating indicators between the proposed method and the improved method at each threshold. Among the results, F1 values higher than 0.8 are marked in pink, which indicates that the entire mapping image achieved satisfactory results at the corresponding threshold. Based on Figure 11, the pink area corresponding to our method is larger than the improved method. Therefore, this shows that our road network extraction method has a stronger ability to capture the colour features between the road network and other features than other road network extraction methods, as well as stronger robustness due to the weakly supervised road extraction method. Moreover, this method has outstanding pixel adaptability during the processing of mapping image binarisation, with more than a 50-pixel scale (183–241). Compare with the WSGAN, the method of (b) contains a 13-pixel scale (234–247). Thus, our method provides a better solution to the difficulties associated with obtaining accurate pixel values during practical applications.

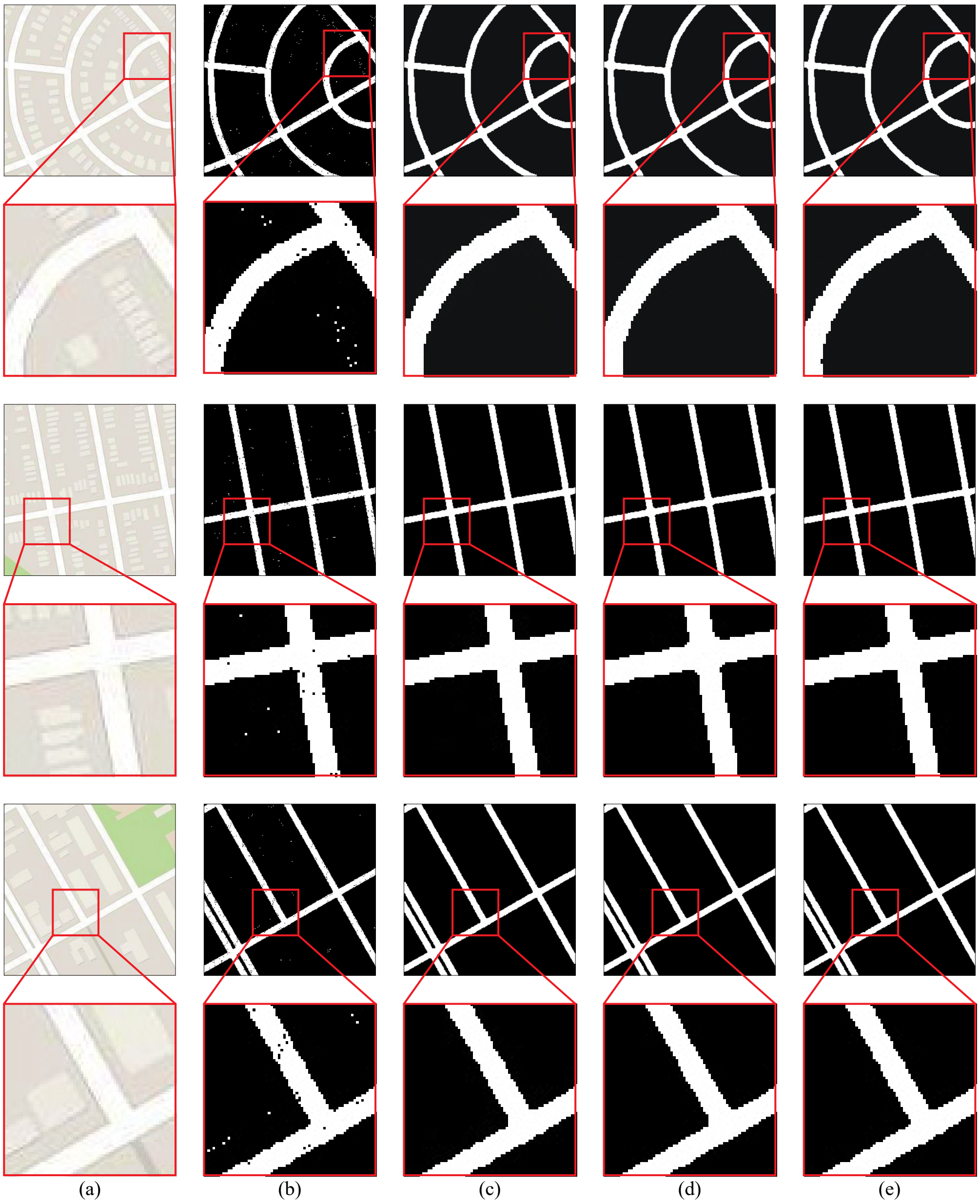

Figure 10. Comparison of several removal results: (**a**) mapping images, (**b**) binary images, (**c**) DE results, (**d**) median filtering results, and (**e**) Gaussian filtering results.

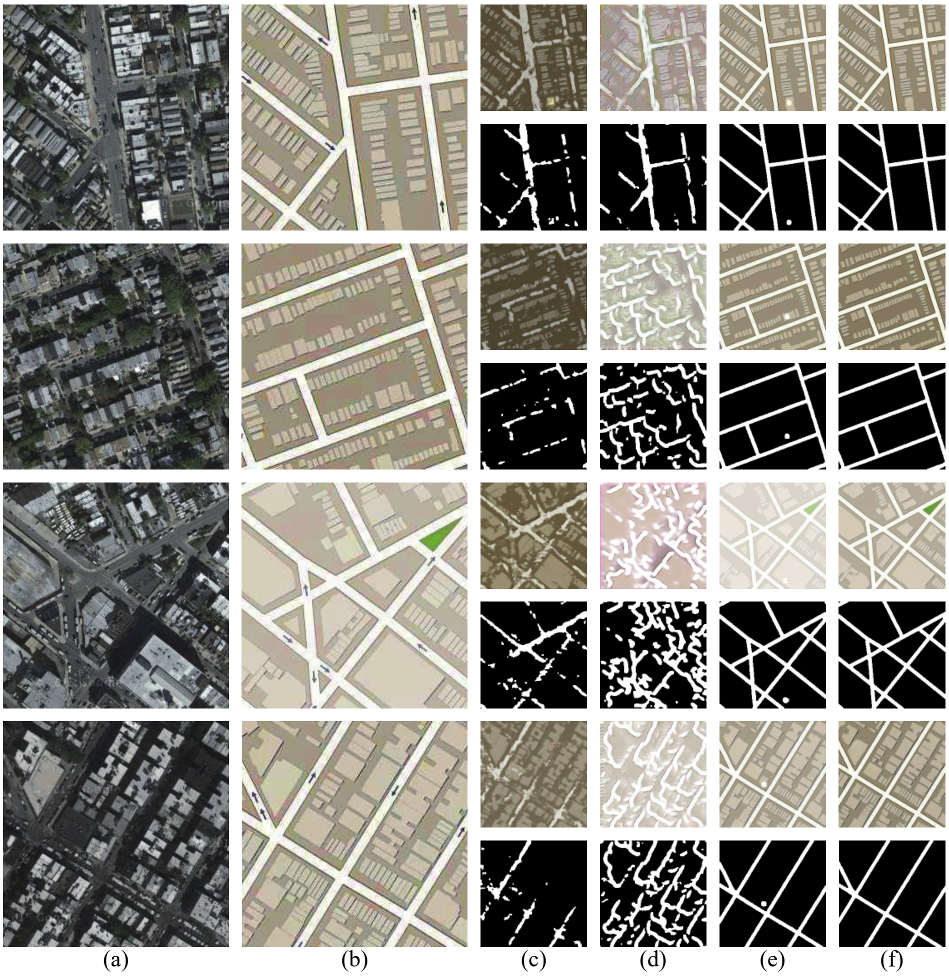

Figure 11. Results of the four mapping networks: (**a**) remote sensing image, (**b**) mapping image, (**c**) the standard GAN with the Res-net framework's generated mapping and binary images, (**d**) the standard GAN's generated mapping and binary images, (**e**) the WGAN with the Res-net network's generated mapping and binary images, and (**f**) the WGAN-GP with the Res-net network's generated mapping and binary images (WSGAN).

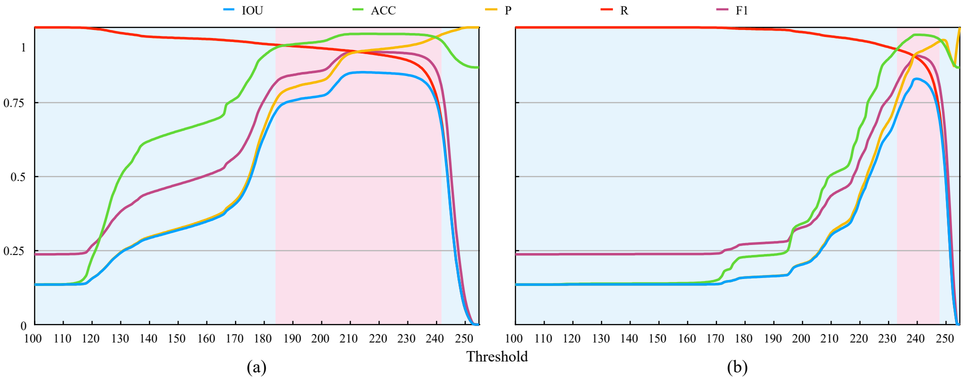

Figure 12. The colour information extraction ability compared with our method: (**a**) our result (WGAN-GP loss and Res-net) and (**b**) the results of the improved method (WGAN loss and Res-net).

To evaluate the binary results through the image visual perception method, Table 1 lists the IoU, ACC, P, R, and F1 scores for the WSGAN method and standard GAN with the Res-net framework to quantitatively evaluate the effectiveness. By comparing the results of the two methods (Table 1), this study can observe that the improved GAN (this study) obtained better results, which is bold the bigger score in the Table 1.

Table 1. The scores of binary results (bold denotes the high values).

Method (Threshold)	IoU	ACC	P	R	F1
This study (214)	**0.849**	**0.978**	**0.920**	**0.917**	**0.918**
WGAN + Res-net (240)	0.827	0.975	0.913	0.898	0.905

The standard GAN with the Res-net framework is an outstanding image transfer method, which used numerous Res-net blocks in the generation network to promote the ability of high-frequency information extraction. However, this method still does not solve the problem associated with the standard GAN training process. Benefiting from the WGAN-GP loss, the WSGAN model achieved relatively satisfactory performance with respect to the IoU, P, R, ACC, and F1 scores; therefore, the WSGAN model generally yields excellent performance.

Benefiting from the Res-net block and GAN structure, the proposed WSGAN model achieved outstanding performance in complex remote sensing scenes. At the same time, to clearly display the road network results, this study used several colours to represent the FP (False Positive), FN (False Negative), road network, and background, where the green colour is FP, red is FN, white is the road network, and black is background. The results of the analysis are shown in Figure 13.

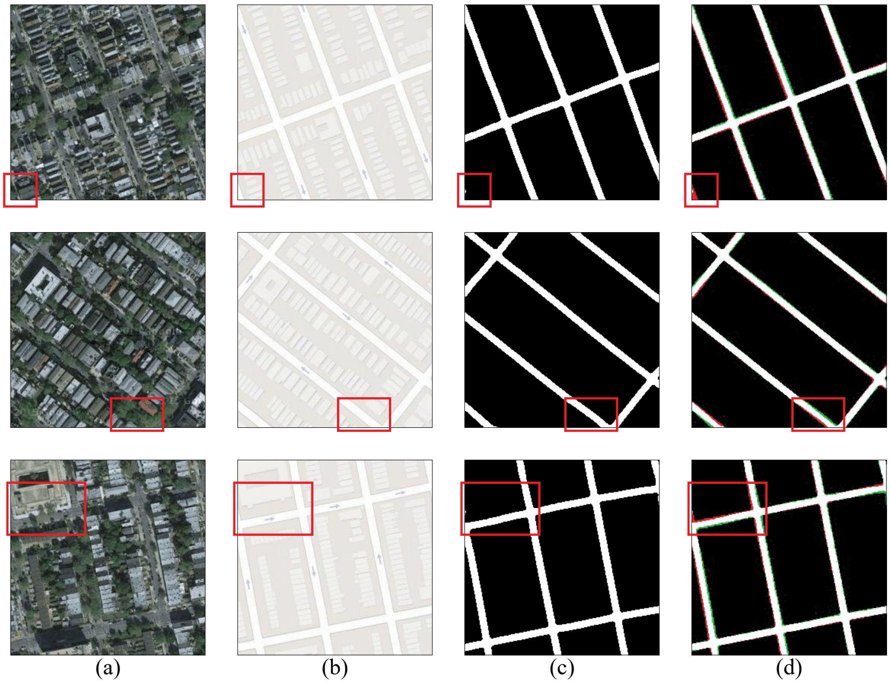

Figure 13. The analysis about the FP and FN: (**a**) remote sensing image, (**b**) mapping image, (**c**) the GT's binary image, and (**d**) synthesis binary image.

From Figure 13, we can see that the road network has great performance in the shelter region and the false-positive area, and false-negative areas are very few. Especially, the shape of the road is also restored, which is total occlusion by the tree. However, there still exists some error in those results; the region in no prior information (such as the first row) has too thick trees, and has building boundary information (such as the third row) that is very difficult to restore.

5. Conclusions

In this study, a weakly supervised method was designed for road network extraction from remote sensing images, termed the WSGAN. The proposed method can be trained by an unpaired dataset with three steps. (1): The mapping network transforms the style of the remote sensing image to obtain the mapping image, which removes textural details and gives several colours for the ground object. The mapping network uses the GAN structure, which can greatly avoid artificial errors and automatically fits the generator from remote sensing images and mapping images. Through the auto discriminate in D, this study avoided design loss function by subjective experience, and the generated mapping image enhanced the road network's feature. The generator used the Res-net block to extract more high-frequency ground object information, and in the optimisation loss function, the WGAN-GP loss solves the collapse problem. (2): To extract the road networks, the mapping image was binarised by a threshold value. To obtain the optimal binary threshold, the evaluation indices were estimated at each threshold, where 213 was the optimal value. (3): As the mapping network is not a semantic segmentation model, the binary image was unavoidably covered with salt-and-pepper noise. To solve this problem, this study selected the DE method, which effectively removed salt-and-pepper noise and retained the original information.

General learning methods for extracting roads from remote sensing images require a significant number of paired samples, where every pixel must be labelled as a road or not. In this study, a weakly supervised framework based on the GAN, which can be trained by easily obtained images (remote sensing images and corresponding mapping images), was designed. This method yielded outstanding mapped road network results, even though some roads were covered by the shadows of buildings or trees. In binarised processing, the experimental results show outstanding performance in obtaining a binary road network image, with the extraction of clear edges and a straight road network. In the future, this study will focus on cartography and use the proposed method to extract other terrain features. To obtain more accurate results, we will further optimise the mapping model. Due to unique training, the weakly supervised method cannot be evaluated similarly to other supervised methods. In future research, we will focus on establishing evaluation indices for the weakly supervised method. Especially, time complexity is also an important study point in after-years, and we will try to optimise network structure and reduce the number of parameters.

Author Contributions: A.H. and Y.X. proposed the road extract architecture design. A.H. and Q.Q. performed the experiments and analysed the data. A.H. wrote the paper. S.C. proposed the figure in this paper. Z.X. and L.W. revised the paper and provided valuable advice for the experiments. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by the National Natural Science Foundation of China (42001340, 41871311, 41871305), the Key Laboratory of Geological Survey and Evaluation of Ministry of Education (GLAB2020ZR05), and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (KF-2020-05-068).

Data Availability Statement: This study did not report any data.

Acknowledgments: The authors thank Siqiong Chen for helping to improve figures.

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

References

1. Zhao, W.; Persello, C.; Stein, A. Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 119–131. [CrossRef]
2. Xu, Y.; Xie, Z.; Wu, L.; Chen, Z. Multilane roads extracted from the OpenStreetMap urban road network using random forests. *Trans. GIS* **2018**, *23*, 224–240. [CrossRef]
3. Liu, H.; Li, H.; Rodgers, M.O.; Guensler, R. Development of road grade data using the United States geological survey digital elevation model. *Transp. Res. Part C Emerg. Technol.* **2018**, *92*, 243–257. [CrossRef]

4. Wu, Z.H.; Zhou, C.J.; Huang, X.L. Main active faults and seismic activity along the Yangtze River Economic Belt: Based on remote sensing geological survey. *China Geol.* **2020**, *3*, 314–338.
5. Roberts, C.W.; Pierce, B.L.; Braden, A.W.; Lopez, R.R.; Silvy, N.J.; Frank, P.A.; Ransom, J.D. Comparison of Camera and Road Survey Estimates for White-Tailed Deer. *J. Wildl. Manag.* **2006**, *70*, 263–267. [CrossRef]
6. Liu, J.; Duan, Y. Saliva: A potential media for disease diagnostics and monitoring. *Oral Oncol.* **2012**, *48*, 569–577. [CrossRef]
7. Huang, L.; Liu, W.; Huang, W. Remote sensing monitoring of winter wheat powdery mildew based on wavelet analysis and support vector machine. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 188–195.
8. Sreekala, B.; John, R.; Johnny, M. Soybean Disease Monitoring with Leaf Reflectance. *Remote Sens.* **2017**, *9*, 127.
9. Guo, R.; Hassan, A.; Hu, Y. Road traffic accident data collection and analysis for road safety research. *Proc. Infants* **2005**, *22*, 134–136.
10. Dupont, E.; Papadimitriou, E.; Martensen, H.; Yannis, G. Multilevel analysis in road safety research. *Accid. Anal. Prev.* **2013**, *60*, 402–411. [CrossRef]
11. Zhu, Y.S. Finite element method research on road slide stability analysis. *J. Highw. Transp. Res. Dev.* **2007**, *24*, 39–42.
12. Huang, S.-X.; Li, S.-Y.; Zhang, X.-G.; Kong, B.; Zhu, Y.-L.; Liu, K.-L. Epidemiological research and analysis on the impaired person in road traffic accident in Chengdu area. *Fa Yi Xue Za Zhi* **2007**, *23*, 269–273. [PubMed]
13. Pereira, L.G.; Janssen, L. Suitability of laser data for DTM generation: A case study in the context of road planning and design. *ISPRS J. Photogramm. Remote Sens.* **1999**, *54*, 244–253. [CrossRef]
14. Xu, Y.; Xie, Z.; Chen, Z.; Xie, M. Measuring the similarity between multipolygons using convex hulls and position graphs. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 847–868. [CrossRef]
15. Adamatzky, A.; Jones, J. Road Planning with Slime Mould: If Physarum Built Motorways it Would Route M6/M74 Through Newcastle. *Int. J. Bifurc. Chaos* **2010**, *20*, 3065–3084. [CrossRef]
16. Skibniewski, M.; Hendrickson, C. Automation and Robotics for Road Construction and Maintenance. *J. Transp. Eng.* **1990**, *116*, 261–271. [CrossRef]
17. Jullien, A.; Dauvergne, M.; Cerezo, V. Environmental assessment of road construction and maintenance policies using LCA. *Transp. Res. Part D Transp. Environ.* **2014**, *29*, 56–65. [CrossRef]
18. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F.; et al. Geographic Object-Based Image Analysis–Towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [CrossRef]
19. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
20. Shi, W.; Miao, Z.; Wang, Q.; Zhang, H. Spectral–Spatial Classification and Shape Features for Urban Road Centerline Extraction. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 788–792. [CrossRef]
21. Zhang, Q.; Couloigner, I. Accurate Centerline Detection and Line Width Estimation of Thick Lines Using the Radon Transform. *IEEE Trans. Image Process.* **2007**, *16*, 310–316. [CrossRef]
22. Zang, Y.; Wang, C.; Cao, L.; Yu, Y.; Li, J. Road Network Extraction via Aperiodic Directional Structure Measurement. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3322–3335. [CrossRef]
23. Mayer, H. Automatic road extraction based on multi-scale modeling, context, and snakes. *Int. Arch. Photogramm. Remote Sens.* **1997**, *32*, 106–113.
24. Song, M.; Civco, D. Road Extraction Using SVM and Image Segmentation. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 1365–1371. [CrossRef]
25. 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]
26. Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A Higher-Order CRF Model for Road Network Extraction. Available online: https://ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/cvpr2013_1227_cr.pdf (accessed on 23 June 2021).
27. Xu, Y.; Xie, Z.; Feng, Y.; Chen, Z. Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sens.* **2018**, *10*, 1461. [CrossRef]
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
29. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H.S. Conditional Random Fields as Recurrent Neural Networks. Available online: https://www.robots.ox.ac.uk/~{}szheng/papers/CRFasRNN.pdf (accessed on 23 June 2021).
30. 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**. [CrossRef]
31. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. Available online: https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html (accessed on 23 June 2021).
32. Rezaei, M.; Harmuth, K.; Gierke, W.; Kellermeier, T.; Fischer, M.; Yang, H.; Meinel, C. A Conditional Adversarial Network for Semantic Segmentation of Brain Tumor. Available online: https://arxiv.org/abs/1708.05227 (accessed on 23 June 2021).

33. Pan, X.; Zhao, J.; Xu, J. Conditional Generative Adversarial Network-Based Training Sample Set Improvement Model for the Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–17. [CrossRef]
34. Hu, A.; Xie, Z.; Xu, Y.; Xie, M.; Wu, L.; Qiu, Q. Unsupervised Haze Removal for High-Resolution Optical Remote-Sensing Images Based on Improved Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 4162. [CrossRef]
35. Xu, M.; Li, Y.; Zhong, J.; Zhang, Y.; Liu, X. Edge Prediction Net for Reconstructing Road Labels Contaminated by Clouds. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa Village, HI, USA, 26 Septembe–2 October 2020; pp. 6969–6972.
36. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein gans. Advances in neural information processing systems. *arXiv* **2017**, arXiv:1704.00028.
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
38. Gil, J.Y.; Kimmel, R. Efficient dilation, erosion, opening, and closing algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 1606–1617. [CrossRef]
39. Liang, L.; Deng, S.; Gueguen, L.; Wei, M.; Wu, X.; Qin, J. Convolutional neural network with median layers for denoising salt-and-pepper contaminations. *Neurocomputing* **2021**, *442*, 26–35. [CrossRef]
40. Li, C.; Wand, M. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision, Cham, Switzerland, 8–16 October 2016; pp. 702–716.
41. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976.
42. Van Erven, T.; Harremos, P. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]
43. Grosse, I.; Galván, P.; Ángel, B.; Carpena, P.; Román-Roldán, R.; Oliver, J.L.; Stanley, H.E. Analysis of symbolic sequences using the Jensen-Shannon divergence. *Phys. Rev. E* **2002**, *65*, 041905. [CrossRef] [PubMed]
44. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv* **2017**, arXiv:1701.07875.
45. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
46. Iizuka, S.; Simo-Serra, E.; Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph.* **2017**, *36*, 1–14. [CrossRef]
47. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative Image Inpainting with Contextual Attention. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514. [CrossRef]