ISPRS Journal of Photogrammetry and Remote Sensing 198 (2023) 169–183

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

SemiRoadExNet: A semi-supervised network for road extraction from remote sensing imagery via adversarial learning

Hao Chen, Zhenghong Li, Jiangjiang Wu, Wei Xiong *, Chun Du

College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, Hunan, China

ARTICLE INFO

Keywords:
Semi-supervised learning
Road extraction
Generative adversarial network
Entropy map
Pseudo-labels

ABSTRACT

Road extraction from remote sensing imagery is a popular and frontier research focus, since road information plays an essential role in application fields, such as urban management, map updating and traffic planning. Deep learning based methods have shown their dominance on road extraction from remote sensing imagery. However, the performance of the existing road extraction methods relies heavily on a large amount of high-quality annotated training data, which is usually hard to obtain in practice. Current semi-supervised road extraction models can effectively reduce the dependency on the labeled data, nevertheless they cannot fully utilize the latent information of low-confidence pixels in pseudo-labels effectively. These pixels are usually a border between road and non-road area and of importance for the prediction accuracy of road extraction models. In order to address these issues, we proposed a novel semi-supervised road extraction network, named SemiRoadExNet. SemiRoadExNet is based on a Generative Adversarial Network (GAN), containing one generator with two discriminators. Firstly, both labeled and unlabeled images are put into the generator network for road extraction, and the outputs of the generator not only include road segmentation results but also the corresponding entropy maps. The entropy maps represent the confidence of prediction (road or non-road) for each pixel. Then, the two discriminators enforce the feature distributions keeping the consistency of road prediction maps and entropy maps between the labeled and unlabeled data. During the adversarial training, the generator is continuously regularized by exploiting the potential information from unlabeled data, thus the generalization capacity of the proposed model can be improved effectively. Compared to several state-of-the-art semi-supervised semantic segmentation methods, the proposed SemiRoadExNet achieves 0.96–5.38% IoU improvements on DeepGlobe Road Extraction, Massachusetts Roads and CHN6-CUG datasets respectively. The source code of d SemiRoadExNet is freely available at https://github.com/hchen118/SemiRoadExNet.

1. Introduction

With the implementation of increasingly growing Earth observation programs, Remote Sensing (RS) images have been developing explosively, and we have entered an age of RS big data (Bong et al., 2009; Li et al., 2016; Chi et al., 2016; Ma et al., 2015). Road extraction from remote sensing imagery has been a popular research topic because of the wide application such as urban management (Bong et al., 2009; Wang et al., 2022a), map updating (Manandhar et al., 2018, 2019) and traffic planning (Chang et al., 2012; Miyamoto et al., 2019; Peddinti et al., 2021). However, road extraction is also one of the most challenging problems, due to the long and thin shape as well as the shades and occlusions induced by vegetation and buildings (Shamsolmoali et al., 2020; Wang et al., 2022b). Besides, the pixels of roads are much lesser than that of non-roads in remote sensing imagery, which leads to the imbalance learning issue for road extraction (Tao et al., 2019).

Recently, deep learning technology has been dominant in the road extraction problem for remote sensing imagery (Abdollahi et al., 2020; Chen et al., 2022a; Lu et al., 2022). Compare to conventional road extraction methods using hand-crafted features (Miao et al., 2012; He et al., 2012), the performance of deep learning-based road extraction models is usually better (Lian et al., 2020). However, the superior performance of these deep learning-based methods highly depends on massive quantities of precisely annotated training data (Li et al., 2021a). Comparing to abundant RS images, labeled RS images are insufficient. A report (Cordts et al., 2016) indicates that pixel-level annotations of one Cityscapes image take almost 90 min on average. Another report (Demir et al., 2018) also notes that the pixel-level annotations of DeepGlobe dataset need professional annotators but human errors are inevitable. Due to complex structures (Yue et al.,

* Corresponding author.
E-mail addresses: hchen@nudt.edu.cn (H. Chen), lizhenghong21@nudt.edu.cn (Z. Li), wujiangjiang@nudt.edu.cn (J. Wu), xiongwei@nudt.edu.cn (W. Xiong), duchun@nudt.edu.cn (C. Du).

https://doi.org/10.1016/j.isprsjprs.2023.03.012
Received 28 October 2022; Received in revised form 10 March 2023; Accepted 10 March 2023
Available online 17 March 2023

2019) and inter-class confusion, pixel-level annotations for RS images are always labor-intensive and time-consuming. Therefore, the road extraction models based on deep learning usually suffer from either insufficient training data or high costs of manual annotation.

To address the problem of lacking enough RS images with accurate road annotations, researchers mainly proposed weakly-supervised methods, unsupervised methods and semi-supervised methods (Chen et al., 2022a; Liu et al., 2022).

The weakly-supervised methods only utilize scribble annotations (such as bounding boxes, scribbles, points and image-level labels) for each category instead of high-quality full annotation data. Wu et al. (2019) and Chen et al. (2022b) proposed novel weakly-supervised road network extraction methods using the OSM (OpenStreetMap) road centerline as weakly-supervised information. Lian and Huang (2021) designed a weakly-supervised road segmentation method adopting points annotations as weakly labels. Hu et al. (2021) and Wei and Ji (2021) proposed a weakly-supervised road extraction method leveraging mapping images as extra scribble road annotations. Although scribble road annotations are easier to be obtained than pixel-wise road annotations, weakly-supervised learning still requires much manual annotations and complicated interactions by human.

The unsupervised road extraction methods do not need annotated training samples and usually adopt clustering algorithms and generative algorithms. Miao et al. (2014) proposed an unsupervised road extraction approach based on mean shift clustering. Tao et al. (2017) utilized an unsupervised-restricted deconvolutional neural network (URDNN) to process remote sensing image classification. URDNN can effectively solve the problem that the existing training datasets need much manual annotations by learning end-to-end and pixel-to-pixel classification. Song et al. (2021, 2022) designed an unsupervised domain mapping model based on adversarial learning called MapGen-GAN, which can quickly transform remote sensing imagery into general maps end-to-end, and the road extraction can be easily carried out from the generated general maps. However, the accuracy of these unsupervised road extraction methods is generally lower than that of supervised or semi-supervised road extraction methods (Liu et al., 2022).

The semi-supervised learning methods leverage a small set of labeled data and a large amount of unlabeled data. They can learn potential patterns from unlabeled samples and regularize the neural network, alleviating the dependency on labels (Sun et al., 2022). The accuracy of semi-supervised learning methods can even be very close to that of full-supervised methods, if the amount of unlabeled data is very large and the model can exploit the potential structural information from both the labeled and unlabeled data effectively (Chen et al., 2022a). Hung et al. (2019) and Zheng et al. (2022) utilized adversarial learning structure for semi-supervised semantic segmentation task. They designed a discriminator to differentiate the predicted probability maps from the ground truth segmentation distribution. Mittal et al. (2019) proposed a two-branch semi-supervised semantic segmentation network to enables low-level and high-level consistency based on generative adversarial network. Li et al. (2021b) and Yang et al. (2022) introduced self-training manner to semantic segmentation network to generate high-quality pseudo labels for unlabeled samples and leverage them with a well-designed strategy.

The core challenge in semi-supervised setting lies in how to extract potential information from unlabeled data effectively. The quality and use strategy of pseudo labels play an essential role in semi-supervised methods. Whereas, incorrect (usually low-confidence) pseudo-labels can degrade the performance of a semi-supervised model dramatically. CPS (Chen et al., 2021b) employ two networks with the same structure and different initialization to conduct semi-supervised semantic segmentation training, when incorrect pseudo-labels are adopted to supervise the training of any one network, the prediction accuracy of both two networks will get worse (You et al., 2022). Therefore, some semi-supervised learning method such as FMWDCT (You et al., 2022) and FixMatch (Sohn et al., 2020) set a confidence threshold to filter samples with low-confidence pseudo-labels.

For road extraction task, the potential information contained in low-confidence pixels of pseudo-labels are of importance for the prediction accuracy of the model, since these pixels are usually in the border between road and non-road area, shown as Fig. 1. Fig. 1(a) is the original RS image. Fig. 1(b) is the road prediction image (white area means road area, black area means non-road area). Fig. 1(c) is the entropy map of the road prediction result, and the entropy value of road border is evidently higher. In Fig. 1(c), the entropy map roughly outlines road and can provide shape information. In Fig. 1(d), high confidence area is marked by red (road) or black (non-road), low-confidence area is marked by green. The latent information of the pixels in green area is usually unused because its high-uncertain pseudo label may mislead network training. If we simply filter the low-confidence pixels out(which is widely adopt in semi-supervised semantic segmentation (Hung et al., 2019; Mittal et al., 2019; You et al., 2022)), the latent information of them is underutilized.

In order to address these issues, we propose a novel GAN based semi-supervised road extraction network named SemiRoadExNet, which consists of one generator and two discriminators. The generator of SemiRoadExNet not only outputs road extraction results but also the corresponding pixel-wise entropy maps. The entropy maps represent the confidence (the value of road prediction) of the prediction for each pixel. The two discriminators try to distinguish both the road extraction results and the corresponding entropy maps from which dataset, labeled or unlabeled. This mechanism can be regarded as a form of consistency regularization in semi-supervised learning. Therefore, the proposed model can utilize the potential information from high-confident pixels to low-confidence counterparts in pseudo labels effectively.

The contributions of this paper are summarized as:

1. A novel framework named SemiRoadExNet is proposed for semi-supervised road extraction based on GAN. The framework extracts structural information from both labeled and unlabeled data, and attempts to make a full use of the potential information of low-confidence pixels in pseudo labels.
2. In order to address the imbalanced learning problem of road and non-road extracting from remote sensing imagery, an encoder–decoder structure with an attention module, rotation consistency constraint, dilated convolution, and skip connections are designed as the generator of the proposed model. The model is proved to be effective toward complex road structure.
3. To validate the proposed method, we conduct extensive experiments on three real world datasets. Comprehensive comparisons and ablation studies are carried out to show the effectiveness and efficiency of the proposed model. Our model achieves a significant improvement in comparison with several state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 introduced related work of road extraction and semi-supervised learning. The proposed SemiRoadExNet is illustrated in details in Section 3. Experiments and analyses are presented in Section 4. In Section 5, we conduct ablation study and provide a comprehensive discussion. Section 6 drew the conclusions of the paper.

2. Related work

In this section, we will illustrate the concept and methods of road extraction from RS imagery, thereafter we will introduce the research of semi-supervised learning for semantic segmentation.

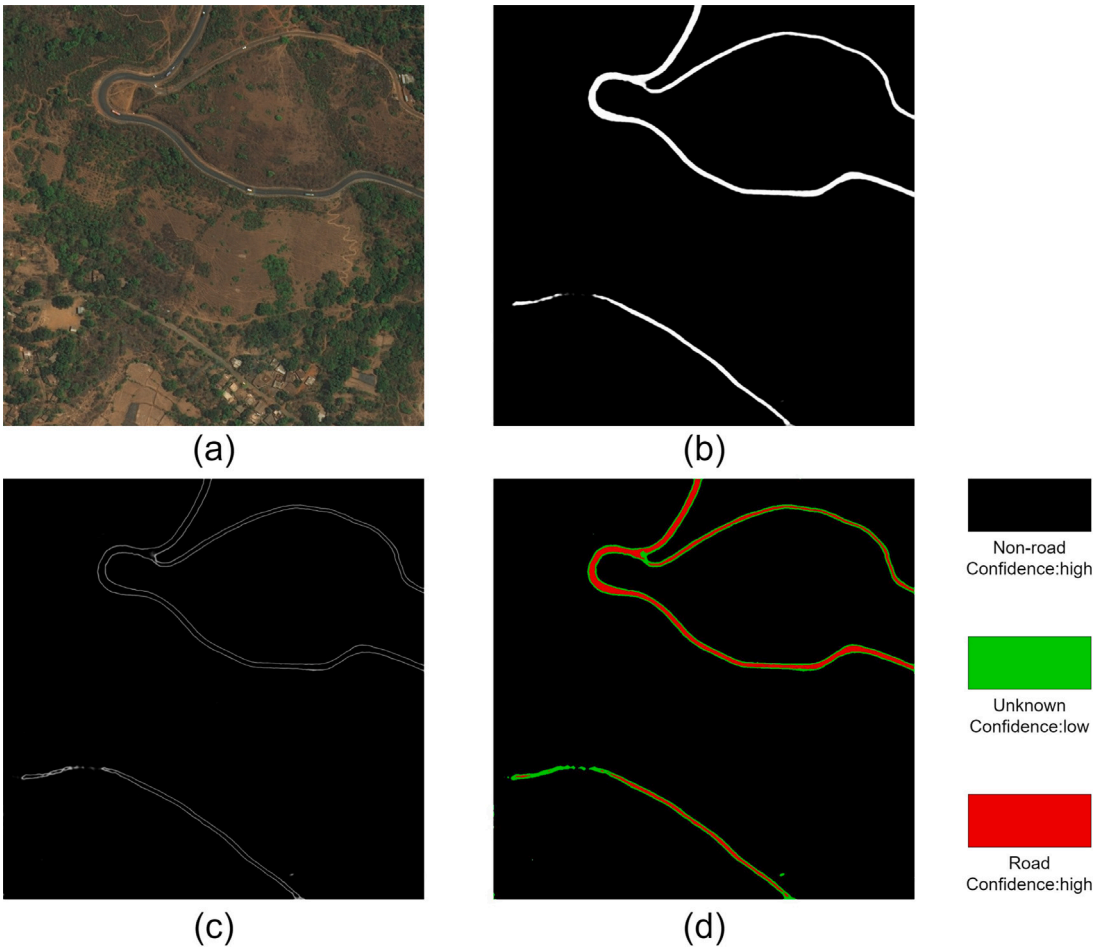

Fig. 1. The low-confidence pixels of a pseudo-label in road extraction task, and they are usually the border between road and non-road area. (a) The original RS image. (b) The road prediction image. (c) The entropy map of the road prediction image. (d) High confidence area and low confidence area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. Road extraction from remote sensing imagery

Road extraction from remote sensing imagery can be seen as a semantic segmentation task. The target of semantic segmentation is assigning semantic labels to image pixels (Chen et al., 2021a). Road extraction is a binary semantic segmentation that assigns road labels and non-road labels to all pixels.

Despite road extraction gains much attention, the task is still challenging. Because of shadows and roadside objects (such as trees, buildings and cars), obscured road regions are easily assigned nonroad labels (Xu et al., 2021). The features of other nonroad regions (such as car parking) are similar to road regions in RS images, therefore these nonroad regions are easily assigned road labels (Abdollahi et al., 2018). Besides, the shape of road is thin and long, which leads to the imbalance of road labels and non-road labels (Wu et al., 2019)(He et al., 2019).

Road extraction methods can be divided into heuristic methods and machine learning methods. Heuristic road extraction methods are mainly based on manual selected texture features (Sghaier and Lepage, 2015), spectral features (Wang et al., 2014), geometric features (He et al., 2012), etc. Heuristic methods discard the dependency on labeled data yet, they are typically time-consuming and error-prone (Liu et al., 2022). Conventional machine learning based road extraction methods mainly include Support Vector Machine (SVM) (Song and Civco, 2004), Artificial Neural Network (ANN) (Kirthika and Mookambiga, 2011), Markov Random Fields (MRFs) classifier (Wang and Luo, 2005), Maximum Likelihood (ML) classifier (Zhou et al., 2006), Mean shift (Miao et al., 2014), Graph theory (Alshehhi and Marpu, 2017) and so on. They adopt hand-crafted features to predict road and non-road area from remote sensing imagery. As the number of remote sensing images increases and the resolution improves, deep learning models, which can learn and represent latent features automatically, have been leading in road extraction methods for remote sensing images. Performances of deep learning methods are usually better, but massive high-quality annotated data is necessary for training.

Road extraction methods base on deep learning can be mainly divided into four classes (Abdollahi et al., 2020): patch-based CNN (Wei et al., 2017), FCN model (Abdollahi et al., 2019), Deconvolutional Net (Xin et al., 2019) and GANs model (Abdollahi et al., 2021). Patch-based CNN assembles prediction patches to produce final prediction images. FCN model uses the interpolation layer as the final layer to upsample the feature map. Deconvolutional Network contains an encoder for extracting latent features and a decoder for producing prediction results. GANs model uses a generator and a discriminator to generate high-quality road segmentations via an adversarial learning manner. The proposed model SemiRoadExNet is based on GANs, and the generator of the proposed road extraction network adopts Deconvolutional Network as backbone.

2.2. Semi-supervised learning for semantic segmentation

Semi-supervised learning (SSL) was proposed to utilize the information of unlabeled data so as to improve the performance. Usually, the semi-supervised dataset X can be divided into labeled part $X_{\mathcal{L}}$ and

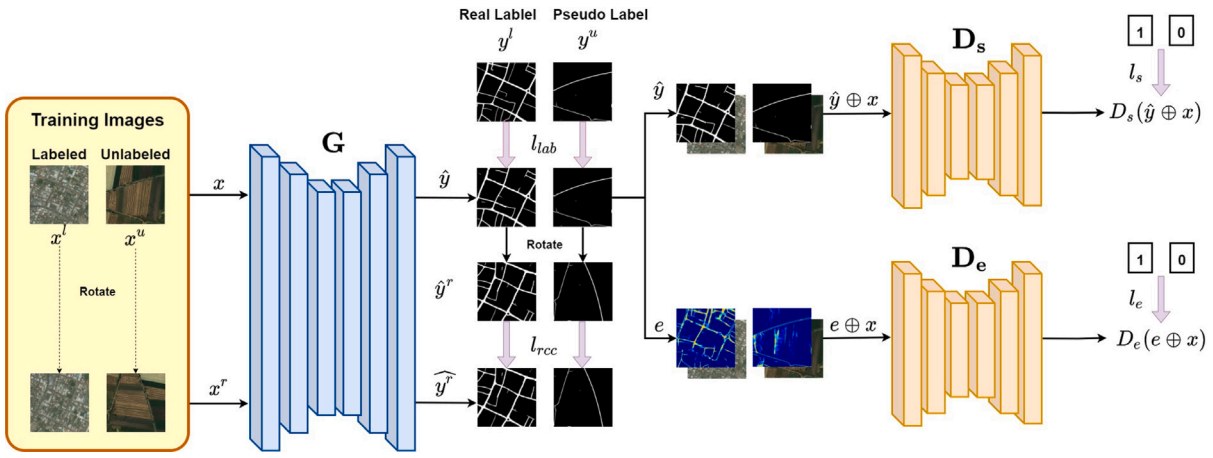

Fig. 2. Overview of proposed SemiRoadExNet. Training images x are sent to generator G to generate road segmentation images $\hat{y}$. Rotated training images x^r are also sent to generator G to generate road segmentation images $\hat{y}^r$. Rotating $\hat{y}$ with the same angle as rotating x thus generates $\hat{y}^r$. Rotation consistency constraint is applied based on $\hat{y}^r$ and rotated $\hat{y}$ (i.e., $\widehat{y^r}$). Road segmentation images with remote sensing images are sent to discriminator D_s to generate pixel-wise image class prediction $D_s(\hat{y} \oplus x)$. The entropy maps with remote sensing images are sent to discriminator D_e to generate pixel-wise image class prediction $D_s(e \oplus x)$.

unlabeled part $X_{\mathcal{U}}$. The labeled part $X_{\mathcal{L}}$ contains M labeled images x_i^l and corresponding labels y_i^l. The unlabeled part contains N unlabeled images x_i^u. Usually, the number of unlabeled images is much larger than that of the labeled images, i.e., $N \gg M$. Fora given network F, its parameters θ_F are optimized by follow method:

$$\min_{\theta_F} \left\{ \frac{1}{M} \sum_{i=1}^{M} \mathcal{L}_s(x_i^l, y_i^l) + \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_u(x_i^u, y_i^u) \right\} \tag{1}$$

where $\mathcal{L}_s$ is supervised loss calculated by labeled data, $\mathcal{L}_u$ is unsupervised loss calculated by unlabeled data.

Semi-supervised learning methods for semantic segmentation can be divided into five classes (Van Engelen and Hoos, 2020): generative methods (Hung et al., 2019; Zheng et al., 2022; Desai and Ghose, 2022), consistency regularization methods (Zhang et al., 2020; He et al., 2022), pseudo-labeling methods (Desai and Ghose, 2022; You et al., 2022; Zou et al., 2020), self-training methods (Li et al., 2021b; Yang et al., 2022; Chen et al., 2021a) and hybrid methods (Wang et al., 2020a, 2021; Zhang et al., 2021). Generative methods simulate real data distribution of both labeled and unlabeled data, and generate new data or high-quality pseudo labels by GAN or VAE. Consistency regularization methods utilize unlabeled data to find a smooth manifold based on manifold assumption or smoothness assumption. Pseudo-labeling methods add some high-quality pseudo labels to the training set as labeled data to enhance the performance of the models iteratively. Self-training methods usually contain a teacher network and a student network. The self-training models are commonly regarded as a form of entropy minimization in SSL, since the re-trained student network is supervised with hard labels produced by the teacher network which is trained on labeled data. Hybrid methods combine pseudo-labeling, consistency regularization, entropy minimization or other methods to enhance the overall performance of the model.

It is notable that the incorrect pseudo labels are still prone to accumulate and degrade the performance of semi-supervised methods (Yang et al., 2022). Therefore, some research works such as FMWDCT (You et al., 2022) and FixMatch (Sohn et al., 2020) set a confidence threshold to filter low-confidence samples. Nevertheless, for the existing semi-supervised methods of remote sensing imagery semantic segmentation and road extraction, the potential information of low-confidence pixels in pseudo-labels is underutilized, since their values are usually below pseudo-label confidence threshold. While, these pixels are usually a border between road and non-road area, and play an essential role in prediction accuracy of the semi-supervised model, shown as Fig. 1.

The proposed SemiRoadExNet belongs to generative method and strives to extract and utilize the potential information from high-confident pixels to low-confidence ones in pseudo labels.

3. Method

In this section, the architecture of proposed SemiRoadExNet will be illustrated firstly. Then, we will present details of the generator and discriminators in SemiRoadExNet. After that, the loss functions will be defined.

3.1. Architecture

The architecture of proposed SemiRoadExNet is illustrated in Fig. 2. Unlike typical architecture of GAN, proposed model contains one generator G (the structure of G is shown in Fig. 3), a road segmentation discriminator D_s and an entropy map discriminator D_e.

The training image set X can be divided into two parts: labeled image set x^l and unlabeled image set x^u. Both of labeled image set and unlabeled image set are feed into the generator G for training. The generator G is a semantic segmentation network and it generates road segmentation images $\hat{y}$(i.e., $G(x)$).

Firstly, the generator G is trained in a supervised fashion and optimized by label loss l_{lab}, which requires predicted road segmentation images $\hat{y}^l$ (i.e., $G(x^l)$) to be consistent with the ground truth labels y^l.

After that, we employ generator G to extract road for unlabeled image set x^u, and get corresponding road extraction result $\hat{y}^u$ (i.e., $G(x^u)$). The high-confident pseudo labels for pixels are calculated, accumulated and added to the training dataset. The pseudo labels are calculated by

$$y_i^u = \begin{cases} 1 & , \quad if\ G(x_i^u) > T_{plc} \\ 0 & , \quad if\ G(x_i^u) < 1 - T_{plc} \\ None & , \quad otherwise \end{cases} \tag{2}$$

where x_i^u is ith pixel of unlabeled image x^u, and y_i^u is ith pseudo pixel and T_{plc} is pseudo label confidence threshold.

Despite pseudo label is a popular and effective method for semi-supervised learning, incorrect pseudo labels can make negative impact on model optimization (Li et al., 2021b). Sun (Sun et al., 2022) also evaluated that a high pseudo label confidence threshold is beneficial to binary semantic segmentation. In order to make pseudo labels more accurate, the road segmentation discriminator D_s is introduced to the proposed model as a typical GAN architecture. The generator G and discriminator D_s can be optimized by road segmentation loss l_s, which will be illustrated in details in Section 3.3. $\hat{y}$ is generated by the generator G. Considering road predictions of unlabeled images are usually worse than predictions of labeled images, D_s is used to distinguish whether $\hat{y}$ is generated from labeled dataset or unlabeled dataset.

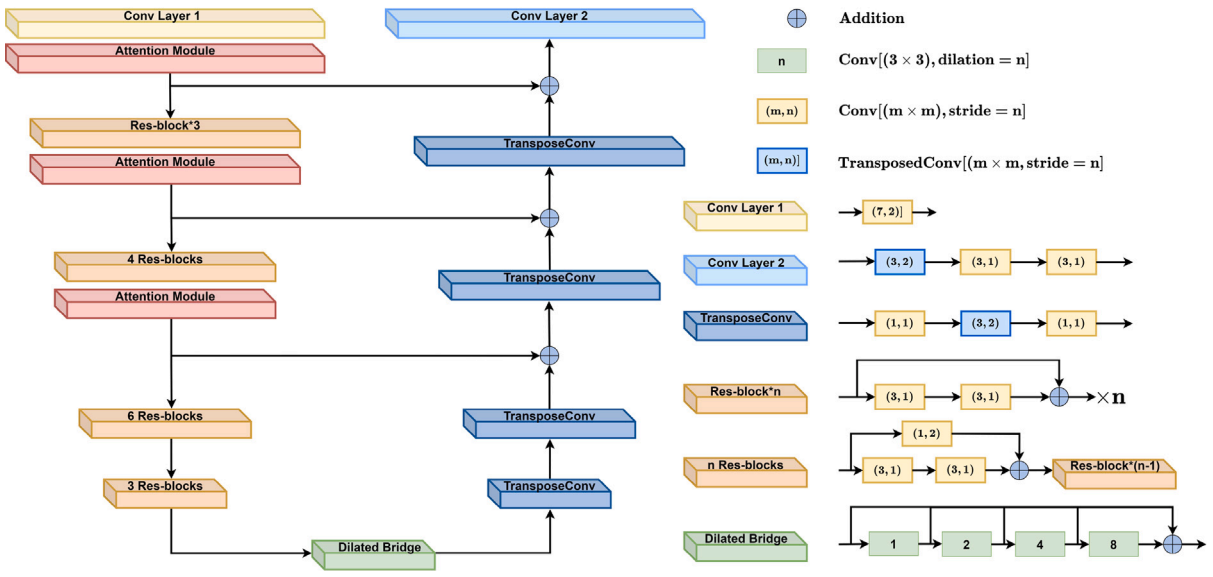

Fig. 3. The structure of generator G.

Nevertheless, unlabeled road segmentation images (i.e., $G(x^u)$) tend to be low-confidence in many areas (Peng et al., 2020), especially in boundary areas. Meanwhile, incorrect pseudo labels are usually with low confidence. Information of low-confidence pixels is usually underutilized in many semi-supervised semantic segmentation methods, since their values are below pseudo label confidence threshold T_{plc}. Besides, some road areas are also low-confidence because of slender shape. In order to utilize the information of low-confidence pixels and relieve label imbalance, we design another branch of discriminator D_e, which is called entropy map discriminator. Considering low-confidence area entropy value of road segmentation image is high while high-confidence area entropy value is low, we extract entropy maps e from road segmentation images $\hat{y}$. Discriminator D_e is designed to distinguish whether the entropy maps e is generated from labeled dataset or unlabeled dataset. And this mechanism can be regarded as another type of consistency regularization between labeled data and unlabeled data. The generator G and discriminator D_e can be optimized by entropy map loss l_e, which will be depicted in detail in Section 3.3.

Moreover, we incorporate the rotation consistency constraint loss l_{rcc} (details in Section 3.3) to further improve the performance of the proposed model. Randomly rotate images x to get rotated images x^r. Randomly rotate road segmentation images $\hat{y}$ to get rotated road segmentation images $\hat{y}$. Rotation consistency constraint requires G having property of rotation consistency. In other words, $\hat{y}^r$ (i.e., $G(x^r)$) should be close to $\hat{y}^r$ (i.e., $G(x)^r$). The generator G can be optimized by rotation consistency constraint loss l_{rcc}, which is based on the consistence of $\hat{y}^r$ and $\hat{y}^r$.

3.2. Generator and discriminators

Like majority GAN models in semi-supervised fashion, the generator G is used to generate semantic segmentation images and the corresponding entropy maps, and the discriminators D_s and D_e are used to distinguish input data labeled or unlabeled.

The structure of the generator G is shown in Fig. 3. The generator takes advantages of skip connections, residual blocks, encoder–decoder architecture and attention mechanism. The generator can be split into three parts: encoder, dilated bridge and decoder. The encoder uses ResNet34 pretrained on ImageNet dataset with attention module. The dilated bridge uses several dilated convolution layers with skip connections. The dilated convolution layers are conducive to expand receptive field. The skip connections are conducive to capture high-level and low-level information. The decoder uses transposed convolution layers to restore the resolution of feature maps.

The attention module is based on mixed domain, which is shown in Fig. 4. It contains a channel attention module which is based on the architecture in (Wang et al., 2020b) and a spatial attention module which is inspired by the coordinate attention mechanism (Huang et al., 2019). The channel attention module is used to effectively extract channel and spatial information so that raises confidence level. The spatial attention module is used to pay more attention to road so that alleviates label imbalance in some extent.

The channel attention module is an extremely lightweight block. The module contains pooling, convolution and sigmoid layers. Height and weight of feature map both reduce to 1 after pooling. Then the convolution operation gives weights to channels. After sigmoid operation, the attention weights are normalized. Finally, the channel attention multiplies with original feature map.

Road is usually thin and long so that the spatial attention module should be non-local to capture long-range information. However, assigning spatial attention values to all pixels is a heavy computation burden with high occupation on GPU memory. Considering road is usually approximately linear, the non-local spatial attention module is designed to assign attention values to vertical direction and horizontal direction. Therefore, the spatial attention module contains vertical module and horizontal module. The feature map passes through vertical module and horizontal module to obtain vertical attention and horizontal attention respectively. Vertical attention adds horizontal attention to obtain non-local spatial attention. The spatial attention multiplies learnable parameter γ, then it adds with original feature map.

The structure of vertical attention module is shown in Fig. 5. The channel number, height and weight are denoted as C, h and w respectively. Feature map generates Q (query), K (key) and V (value) by three different convolution operations. Q dot products K to obtain intermediate feature map with the size of $h \times w \times w$. The intermediate feature passes through softmax operation then dot products V. After aforementioned operations, the size of generated feature map is $h \times C \times w$. Lastly, generated feature map performs dimension transformation to obtain vertical attention. The size of vertical attention is the same as original feature map, i.e., $C \times h \times w$. As Fig. 6 shows, the structure of horizontal attention module is the same as that of vertical attention module. The only difference from vertical attention is the feature map size.

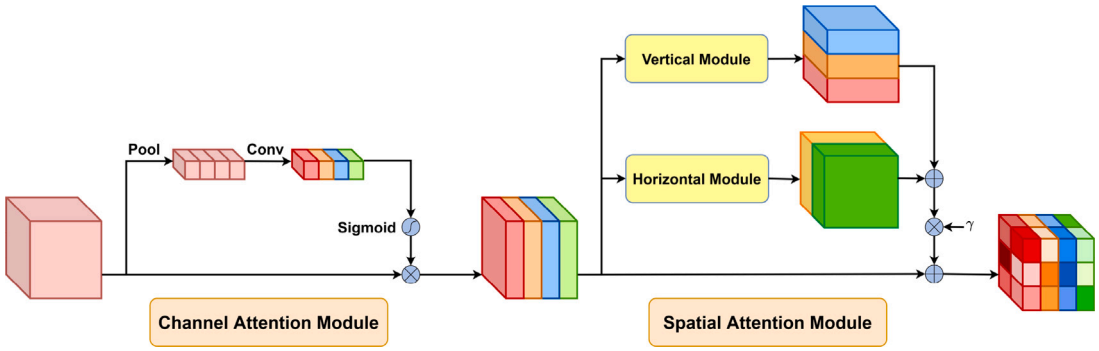

Fig. 4. The structure of attention module.

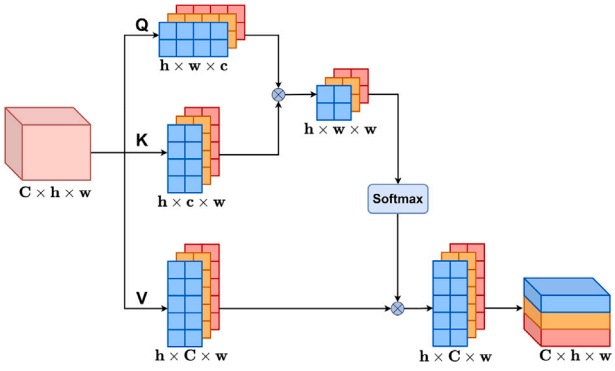

Fig. 5. The structure of vertical attention module.

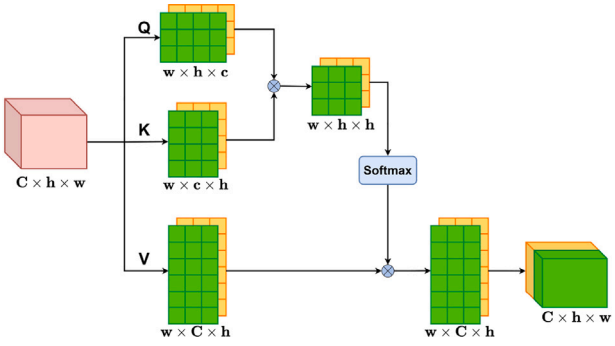

Fig. 6. The structure of horizontal attention module.

The structure of discriminators D_s and D_e employs the UNet (Zhang et al., 2018) as backbone. Unlike the general GAN, the discriminators are both semantic segmentation networks instead of image classification networks. Because the size of remote sensing images is large, the distinguishment difficulty of different areas in an image is different. Adopting semantic segmentation network UNet rather than image classification network as the discriminator can achieve better performance (Zheng et al., 2022). The input of D_s is $\hat{y} \oplus x$, i.e., the concatenation of generated images $\hat{y}$ and corresponding training images x. Similarly, the input of D_e is the concatenation of entropy images e and the corresponding training images x. The inputs of D_s and D_e are both in the size of $4 \times H \times W$. The outputs of discriminators are both possibility maps. The size of possibility map is $1 \times H \times W$. The probability maps represent the pixel-wise possibilities of input data to be from the labeled dataset.

3.3. Loss function

Assume the dataset X consists of: (1) labeled data $x^l = \{(x_i^l, y_i^l)\}_{i=1}^M$, which contains M labeled images x_i^l and corresponding labels y_i^l; (2) unlabeled data $x^u = \{(x_i^u)\}_{i=1}^N$, which contains N unlabeled images x_i^u. Our proposed method contains four losses: road segmentation loss l_s, entropy map loss l_e, label loss l_{lab} and rotation consistency constraint loss l_{rcc}. The generator G is optimized by l_s, l_e, l_{lab} and l_{rcc}. The discriminator D_s is optimized by l_s. The discriminator D_e is optimized by l_e.

The total loss l_{total} of the network is defined as:

$$l_{total} = l_s + l_e + l_{lab} + l_{rcc} \tag{3}$$

where l_s, l_e, l_{lab} and l_{rcc} are defined in 3.3.1, 3.3.2, 3.3.3 and 3.3.4 respectively.

Meanwhile, the total loss for optimizing generator G is:

$$l_G = l_{sg} + l_{eg} + l_{lab} + l_{rcc} \tag{4}$$

where l_{sg} and l_{eg} are defined in 3.3.1 and 3.3.2 respectively.

3.3.1. Road segmentation loss

Road segmentation loss l_s includes discriminator loss l_{sd} and generator loss l_{sg}. Unlabeled road segmentation images are usually less close to its real road distributions than labeled road segmentation images, because the pseudo labels are usually dynamic and not always accurate (Li et al., 2021b). Discriminator D_s is employed to figure out whether the input data is from labeled dataset or unlabeled dataset. The generator G is used to fool the discriminator D_s. Generator G and discriminator D_s align the feature distributions of unlabeled generated images $G(x^u)$ and labeled generated images $G(x^l)$.

The segmentation discriminator loss l_{sd} aims to improve the discriminative ability of D_s between $G(x^l)$ and $G(x^u)$. It is defined as:

$$\begin{aligned} l_{sd} = & \frac{1}{M}\sum_{l=1}^{M} l_{bce}(D_s(G(x_i^l) \oplus x_i^l), \mathbf{1}) + \\ & \frac{1}{N}\sum_{l=1}^{N} l_{bce}(D_s(G(x_i^u) \oplus x_i^u), \mathbf{0}) \end{aligned} \tag{5}$$

where $\oplus$ denotes the concatenation operation and l_{bce} is binary cross entropy loss. The general form of binary cross entropy loss l_{bce} is defined as:

$$l_{bce}(y^p, y^t) = \frac{1}{H \times W}\sum_{i=1}^{H \times W} [-y_i^t log(y_i^p) - (1 - y_i^t) log(1 - y_i^p)] \tag{6}$$

where H and W are image height and image weight respectively, y_i^p and y_i^t are prediction of ith pixel and target of ith pixel respectively.

The segmentation generator loss l_{sg} is used to fool the discriminator. It is computed as:

$$l_{sg} = \frac{1}{N}\sum_{i=1}^{N} l_{bce}(D_s(G(x_i^u) \oplus x_i^u), \mathbf{1}) \tag{7}$$

The whole road segmentation loss l_s is consist of l_{sd} and l_{sg}, it is defined as:

$$l_s = l_{sd} + l_{sg} \tag{8}$$

3.3.2. Entropy map loss

Entropy map loss l_e consists of discriminator loss l_{ed} and generator loss l_{eg}. The generator G usually tends to produce low-entropy predictions with high confidence on labeled images, whereas the predictions on unlabeled images are usually high-entropy value with low confidence. The pseudo labels cannot unitize low confidence pixels information effectively due to the confidence threshold T_{plc}. Hence, we extract entropy maps e of generated images. Entropy maps e consists of labeled generated entropy maps e^l and unlabeled generated entropy maps e^u. As for binary semantic segmentation, entropy map e is computed as:

$$e = -\hat{y} \cdot log(\hat{y}) - (1-\hat{y}) \cdot log(1-\hat{y}) \tag{9}$$

Discriminator D_e is designed to distinguish whether the entropy maps e are from labeled image set or unlabeled image set. The entropy discriminator loss l_{ed} is computed as:

$$\begin{aligned} l_{ed} = & \frac{1}{M}\sum_{i=1}^{M} l_{bce}(D_e(G(e_i^l) \oplus x_i^l), \mathbf{1}) + \\ & \frac{1}{N}\sum_{i=1}^{N} l_{bce}(D_e(e_i^u \oplus x_i^u), \mathbf{0}) \end{aligned} \tag{10}$$

Similarly, the generator G is designed to try to fool the discriminator D_e. The entropy generator loss l_{eg} is computed as:

$$l_{eg} = \frac{1}{N}\sum_{i=1}^{N} l_{bce}(D_e(e^u \oplus x^u), \mathbf{1}) \tag{11}$$

The compete entropy map loss l_e is consist of l_{ed} and l_{eg}, it is defined as:

$$l_e = l_{ed} + l_{eg} \tag{12}$$

3.3.3. Label loss

Label loss l_{lab} consists of binary cross-entropy loss l_{bce} and dice coefficient loss l_{dice}. l_{lab} is used to utilize the full-supervised information of labels y^l and pseudo labels y^u. Non-road pixels are much more than road pixels in road extraction task. Towards the label imbalance problem, we use combination of binary cross-entropy loss l_{bce} and dice coefficient loss l_{dice}. $l_{bce}(\hat{y}, y)$ and $l_{dice}(\hat{y}, y)$ are defined as:

$$l_{bce}(\hat{y}, y) = \frac{1}{n}\sum_{i=1}^{n}[-y_i log(\hat{y}_i) - (1-y_i) log(1-\hat{y}_i)] \tag{13}$$

$$l_{dice}(\hat{y}, y) = \frac{1}{n}\sum_{i=1}^{n}(1 - \frac{2 \times |\hat{y} \cap y|}{|\hat{y}| + |y|}) \tag{14}$$

where n is the number of pixels participating in calculation. For labeled images, all pixels of training images and labels are participated in calculation, therefore n is equal to $H \times W$. For unlabeled images, only high-confident pseudo label pixels and corresponding training image pixels are participated in calculation, therefore n is equal to the pixel number of pseudo label pixels. $|\cdot|$ denotes calculating the sum of pixel values. $\hat{y} \cap y$ denotes the intersection of $\hat{y}$ and y.

The complete label loss l_{lab} is defined as:

$$l_{lab} = l_{bce}(\hat{y}, y) + l_{dice}(\hat{y}, y) \tag{15}$$

3.3.4. Rotation consistency constraint loss

Rotation consistency constraint loss l_{rcc} is based on rotation consistency constraint. Rotation consistency constraint depicts the generalized rotational consistency property of the images from the target domain Li et al. (2021a), i.e., road segmentation images ought to rotate the same degree when input images rotate. l_{rcc} is defined as:

$$\begin{aligned} l_{rcc} &= \frac{1}{H \times W}\sum_{i=1}^{H \times W} l_{bce}((\hat{y}^r), \hat{y}^r) \\ &= \frac{1}{H \times W}\sum_{i=1}^{H \times W} l_{bce}(G(x)^r, G(x^r)) \end{aligned} \tag{16}$$

where x^r denotes randomly rotating training images x, $\hat{y}^r$ (i.e., $G(x^r)$) denotes road segmentation images of rotated images x^r, $\hat{y}^r$(i.e., $G(x)^r$) denotes rotating road segmentation images $G(x)$ at the same angle of rotating x.

4. Experiment

In this section, we will introduce the datasets used in the experiment firstly. Then we will introduce comparative methods and evaluation metrics used in the experiment. In Section 4.3, we will give a detailed description of training. In Section 4.4, we will conduct experiments to verify the effectiveness of proposed method. In Section 4.5, we will conduct ablation experiments to verify the effectiveness of each part in proposed model.

4.1. Dataset

To verify the effectiveness of proposed method, we select three road extraction datasets with different road width: DeepGlobe Road Extraction dataset (Demir et al., 2018), Massachusetts Roads dataset (Mnih and Hinton, 2010) and CHN6-CUG dataset (Zhu et al., 2021). The roads in Massachusetts Roads dataset are slenderest and the roads in CHN6-CUG dataset are widest compared to image width.

(1) DeepGlobe Road Extraction dataset: This dataset consists of 8570 satellite images including 6226 satellite images with labels and 2344 satellite images without labels. The dataset covers around 2220 sq. km area with a resolution of 50 cm/pixel. The size of each image is 1024 × 1024. In the experiment, we used 6226 satellite images with labels. These images are randomly divided into 5626 training images, 300 validation images and 300 test images.

(2) Massachusetts Roads dataset: This dataset consists of 1171 satellite images with labels, including 1108 training images, 14 validation images and 49 test images. The dataset covers around 2600 sq. km area with a resolution of 120 cm/pixel. The size of each image is 1500 × 1500. It covers a wide variety of urban, suburban and rural regions.

(3) CHN6-CUG dataset: This dataset contains 4511 labeled images with a resolution of 50 cm/pixel. The size of each image is 512 × 512. The dataset covers six cities with different levels of urbanization, city size, development degree and urban structure. The dataset includes the Chaoyang area of Beijing, the Yangpu District of Shanghai, Wuhan city center, the Nanshan area of Shenzhen, the Shatin area of Hong Kong, and Macao. However, there are quite a number of images not suitable for training, which are all black, large area black and all dark color sea. Some useless images are shown in Fig. 7. In order to eliminate useless images, we washed some images of the dataset. We consider pixel values lower than 20 as useless pixels, and select images that the proportion of useless pixels is lower than 20%. Finally, we get 3065 images after washing. In the experiment, we randomly split 2453 images as training images, 306 images as validation images and 306 images for test images.

The training set image number and labeled rate are denoted as n and λ respectively. Training set is randomly split into labeled images and unlabeled images. The numbers of labeled images and unlabeled images are $\lambda \times n$ and $(1-\lambda) \times n$ respectively. In the experiment, λ is set as 0.05, 0.1 and 0.2. The number of dataset images is shown in Table 1.

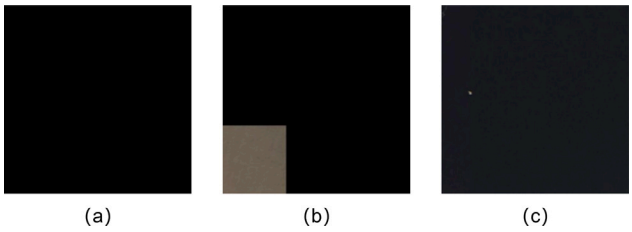

Fig. 7. Examples of washed images in CHN6-CUG dataset. (a) All black. (b) Large area black. (c) All dark sea.

Table 1
The number of dataset images used in the experiment.

Dataset	Training	Validation	Test	Total
DeepGlobe Road Extraction	5626	300	300	6226
Massachusetts Roads	1108	14	49	1171
CHN6-CUG	2453	306	306	3065

4.2. Baselines and evaluation metrics

To verify the effectiveness of proposed SemiRoadExNet, some state-of-the-art methods are selected:

(1) D-LinkNet (Zhou et al., 2018). D-Linknet is a full-supervised road extraction network built with LinkNet architecture. It contains dilated convolution layers to expand receptive field. The network got the best IoU scores in the CVPR DeepGlobe 2018 Road Extraction Challenge.

(2) SII-Net (Tao et al., 2019). SII-Net is a full-supervised road extraction network. SII-Net contains encoder, decoder and SIIS module. The network learns both local characteristics and global spatial structure information. It can preserve the continuity of the extracted road.

(3) RoadExNet. RoadExNet is the generator of proposed SemiRoadExNet. RoadExNet is trained only utilizing the labeled data. Therefore, it is a full-supervised method. We choose RoadExNet as baseline to verify the effectiveness of the architecture of the generator of SemiRoadExNet as a road extraction network.

(4) AdvNet (Hung et al., 2019). AdvNet is a semi-supervised semantic segmentation network. Like the proposed model SemiRoadExNet, AdvNet is also based on GAN, but it only contains one generator and one discriminator. The generator of AdvNet generates semantic segmentation images. And the discriminator differentiates the generated images whether from the ground truth segmentation distribution.

(5) s4GAN (Mittal et al., 2019). s4GAN is a semi-supervised semantic segmentation network containing two branches. The framework links semi-supervised classification with semisupervised segmentation including self-training, which enables low-level and high-level consistency.

(6) ST++ (Yang et al., 2022). ST++ is an advanced semi-supervised semantic segmentation framework. It can progressively leverage the unlabeled images. It has been proved to outperform several previous semi-supervised methods in semantic segmentation tasks.

(7) FMWDCT (You et al., 2022). FMWDCT is the only end-to-end semi-supervised remote sensing road extraction method, which uses consistency regularization and pseudo labels. It contains 4 neural networks for training.

We adopted three evaluation metrics to compare the performance of different methods: F1 score ($F1$), Intersection over Union (IoU) and Kappa coefficient ($Kappa$). The evaluation matrices are formulated as follows:

$$F1 = 2 \times \frac{Pr \times Rec}{Pr + Rec} \tag{17}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{18}$$

$$Kappa = \frac{OA - PRE}{1 - PRE} \tag{19}$$

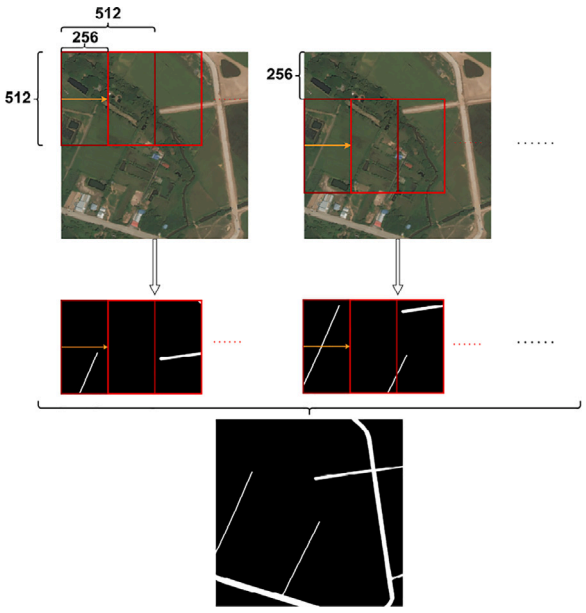

Fig. 8. The process of overlap test.

where TP, FP, TN and FN are the number of true positives, false positives, true negatives and false negatives respectively. Pr, Rec, OA and PRE are defined as

$$Pr = \frac{TP}{TP + FP} \tag{20}$$

$$Rec = \frac{TP}{TP + TN} \tag{21}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

$$PRE = \frac{(TP + FN)(TP + FP) + (TN + FN)(TN + FP)}{(TP + TN + FP + FN)^2} \tag{23}$$

4.3. Implementation details

In the experiment, we use PyTorch as the deep learning framework. The optimizer for the generator and discriminators are Adam optimizers with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, respectively. Learning rate initially is set to 0.004 shared by G, D_s and D_e. A poly-learning policy is utilized to better train the model, where the learning rate decayed to its one fifth every 6 consecutive rises of validation loss. Training process stops when its learning rate smaller than 5×10^{-7}. The pseudo label threshold T_{plc} is set to 0.95 followed as the setting of Sun et al. (2022). Besides, the epoch number and batch size are set as 100 and 4 respectively. The labeled rates for dataset are set as 0.05, 0.1 and 0.2. The code is implemented in Pytorch 1.3 and experiments are run on a server equipped with a NVIDIA 2080 Ti GPU (11 GBs).

Due to the limitation of computing resource and GPU memory, we train and test images with the size of 512 × 512 rather than original size. Since deep learning performance requires a large amount of training data, we make data augmentation through random cropping (image size is 512 × 512 after cropping), random horizontal flip (the probability is 0.5), random vertical flip (the probability is 0.5), random rotation (rotation degrees are random selected from 0, 90, 180 and 270), random shifting (shifting distance is between 0 and 0.1 times the image size), random scaling (image size scales to 0.9–1.1 times) and random color transformation (brightness, contrast, saturation and hue are change by up to 0.05, 0.05, 0.05, 0.1 times respectively). During test period, we test images by overlap test, which is shown in Fig. 8. We move test window from left to right and from up to down. The test window stride gap is 256 pixels. We add the road images generated by the test window and average them to obtain the complete road images.

Table 2
Summary of the quantitative results on DeepGlobe Road Extraction dataset.

Method	5%			10%			20%		
	F1	IoU	Kappa	F1	IoU	Kappa	F1	IoU	Kappa
D-LinkNet Zhou et al. (2018)	0.4646	0.3069	0.4250	0.5022	0.3374	0.4550	0.5094	0.3411	0.4597
SII-Net (Tao et al., 2019)	0.5558	0.3980	0.5283	0.5482	0.3912	0.5225	0.5680	0.4106	0.5449
RoadExNet	0.5768	0.4144	0.5514	0.5551	0.3921	0.5241	0.5971	0.4350	0.5725
AdvNet (Hung et al., 2019)	0.3038	0.1897	0.2885	0.3451	0.2208	0.3294	0.4173	0.2767	0.3993
s4GAN (Mittal et al., 2019)	0.2687	0.1678	0.2557	0.3199	0.2036	0.3046	0.4906	0.3363	0.4674
ST++(Yang et al., 2022)	0.1778	0.1020	0.1600	0.1869	0.1071	0.1685	0.2059	0.1174	0.1820
FMWDCT (You et al., 2022)	0.5853	0.4243	0.5661	0.6328	0.4749	0.6158	0.6745	0.5226	0.6585
SemiRoadExNet	**0.6067**	**0.4514**	**0.5903**	**0.6674**	**0.5145**	**0.6304**	**0.6923**	**0.5443**	**0.6783**

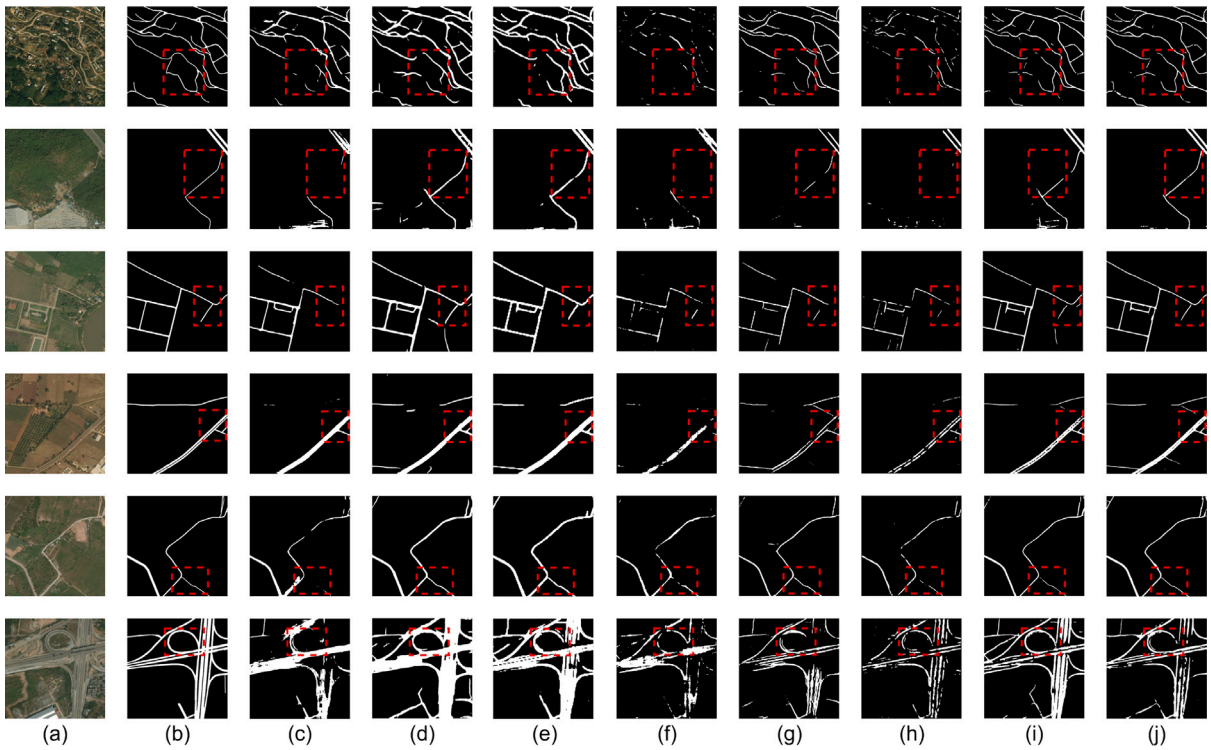

Fig. 9. Visual comparisons of different methods on Deep Globe Road Extraction dataset with 20% labeled rate. (a) Remote sensing image. (b) Ground truth road image. (c) D-LinkNet. (d) SII-Net. (e) RoadExNet. (f) AdvNet. (g) s4GAN. (h) ST++. (i) FMWDCT. (j) SemiRoadExNet.

4.4. Experimental result

To verify the effectiveness of proposed model, extensive experiment results are summarized and analyzed.

4.4.1. DeepGlobe Road Extraction dataset

The quantitative results on DeepGlobe Road Extraction dataset are shown in Table 2. We calculate and summarize F1, IoU and Kappa metrics based on different labeled rates.

From Table 2, it can be seen that the SemiRoadExNet outperforms the other methods in different labeled rate. We can conclude that the proposed SemiRoadExNet achieves the best performance against other methods. SemiRoadExNet gets the highest F1, OA, and Kappa values in the dataset. Meanwhile, RoadExNet gets quite good results. Compared with RoadExNet, SemiRoadExNet gets 3.70%, 12.25% and 10.93% improvements in IoU when labeled rates are 5%, 10% and 20% respectively. SemiRoadExNet gets better performance because it mines unlabeled images information more effectively. Compared with the SII-Net, SemiRoadExNet gets 5.08%, 5.34% and 6.20% improvements of F1, IoU and Kappa respectively, when the labeled rate is 5%. The improvements are 11.92%, 12.33% and 10.79% respectively when the labeled rate is 10%. When the labeled rate increases to 20%, the improvements are 12.43%, 13.37% and 13.34% respectively. The improvements increase with the labeled rate increasing.

Furthermore, SII-Net performances are better than D-LinkNet. SII-Net gets 4.59%, 5.38% and 6.74% improvements of F1, IoU and Kappa respectively, when the labeled rate is 10%. This result demonstrates the advantage of using road-specific contextual information. Note that, AdvNet achieves slightly worse performance than SII-Net, because the generator of AdvNet is not specially designed for road extraction target and it may face mode collapses problems during GAN training (Dai et al., 2017). When the labeled rate is 20%, SII-Net gets 7.74%, 7.43% and 7.75% improvements respectively compared to s4GAN. With regard to s4GAN, its performance is worse than AdvNet when the labeled rate is low. This is because the MLMT branch of s4GAN can mislead the results. Compared to other baselines, performances of ST++ are the worst in F1, IoU and Kappa metrics. This is because ST++ relies pseudo labels much, but pseudo labels of ST++ are not usually reliable enough owing to rare and thin road characteristic.

To see this improvement more visually, we show results of different methods with 20% labeled rate in Fig. 9. From Fig. 9, we can observe that there exist many missed road areas and false road areas in the comparative methods. Whereas the proposed SemiRoadExNet achieves the best visual performance and its results are much closer to the ground truth road images. Specially, roads in the last row are very

Table 3
Summary of the quantitative results on Massachusetts Roads dataset.

Method	5%			10%			20%		
	F1	IoU	Kappa	F1	IoU	Kappa	F1	IoU	Kappa
D-LinkNet Zhou et al. (2018)	0.1121	0.0606	0.0991	0.2116	0.1224	0.1865	0.2429	0.1459	0.2243
SII-Net (Tao et al., 2019)	0.3700	0.2252	0.3296	0.4739	0.3124	0.4424	0.6147	0.4460	0.5915
RoadExNet	0.3865	0.2414	0.3672	0.5065	0.3406	0.4874	0.5456	0.3759	0.5149
AdvNet (Hung et al., 2019)	0.1915	0.1071	0.1616	0.2113	0.1195	0.1853	0.2119	0.1202	0.1901
s4GAN (Mittal et al., 2019)	0.3187	0.1954	0.3039	0.3961	0.2541	0.3793	0.4313	0.2809	0.4132
ST++(Yang et al., 2022)	0.1418	0.0792	0.1353	0.1013	0.0539	0.0945	0.1476	0.0827	0.1406
FMWDCT (You et al., 2022)	0.6475	0.4830	0.6309	0.6883	0.5297	0.6736	0.6883	0.5303	0.6744
SemiRoadExNet	**0.6948**	**0.5362**	**0.6788**	**0.6995**	**0.5445**	**0.6859**	**0.7023**	**0.5466**	**0.6878**

complex. SemiRoadExNet road predictions in the two rows are close to ground truth road images, which proves that SemiRoadExNet adapts to complex road condition.

4.4.2. Massachusetts roads dataset

The quantitative evaluation results of different methods are shown in Table 3. Proposed SemiRoadExNet also gets the highest F1, OA, and Kappa values against other methods in different labeled rate. Meanwhile, proposed RoadExNet gets better results than D-LinkNet and SII-Net. Compared with RoadExNet, SemiRoadExNet gets 29.48%, 20.40% and 17.07% IoU improvements when labeled rates are 5%, 10% and 20% respectively. The improvements are much high in the dataset, so it indicates that SemiRoadExNet can adapt slender road labeled owing to non-local special attention module. When the labeled rate is 20%, SemiRoadExNet gets 8.76%, 10.05% and 9.64% higher F1, IoU and Kappa respectively than SII-Net. The improvements increase to 22.56%, 23.22% and 24.35% respectively in 10% labeled rate. When labeled rate is 5%, the improvements continue increasing to 32.48%, 31.10% and 34.92% respectively. SemiRoadExNet gets higher improvements with less labeled images because it effectively utilizes entropy information of predictions, rotation consistency constraint and pseudo labels of unlabeled data. Compared with FMWDCT, SemiRoadExNet gets 5.31%, 1.48% and 1.63% improvements in IoU when labeled rates are 5%, 10% and 20% respectively. The improvement increases with labeled rate decreasing, which suggests that SemiRoadExNet are better at extract unlabeled information.

Besides, SII-Net performances better than D-LinkNet in the dataset. SII-Net gets 15.05%, 14.26% and 17.31% improvements of F1, IoU and Kappa respectively in 10% labeled rate . This improvement is higher than the improvement in Deep Globe Road Extraction dataset, which indicates the SIIS module is better at capturing slender road information. The performance of s4GAN is slightly worse than SII-Net. The generator of s4GAN is not particularly designed for road extraction, but the unlabeled data information remedies the deficiency. When the labeled rate is 5%, compared to s4GAN, SII-Net only gets 5.13%, 2.97% and 2.57% improvements, respectively. With regard to AdvNet, its performance is worse than s4GAN. This is because Massachusetts Roads dataset labels are road centerline images rather than road area images, which means road proportion is lower, and the MLMT branch of s4GAN plays an important role to filter useless areas.

In order to show performances visually, different methods results with 20% labeled rate in Fig. 10. We can observe that large amount of obliterated road areas and false road areas are in the comparative methods. Whereas proposed SemiRoadExNet achieves the best visual performance and its results are closer to the ground truth road images. As the first row in the figure shows, the road predictions are discontinuous except RoadExNet and SemiRoadExNet. It reflects that RoadExNet and SemiRoadExNet have abilities to overcome roadside obstacles. Note that, the road predictions of ST++ are dark, because the pseudo labels tend to be large-area dark and the pseudo labels influence the result much. The road predictions in SII-Net and RoadExNet are coarse. This is because SIIS module of SII-Net and attention module of RoadExNet both focus on horizontal and vertical information, which means increasing road prediction width is beneficial to decrease punishment. Otherwise, the road predictions of SemiRoadExNet are not as road predictions of RoadExNet. This is because horizontal and vertical confidence of SemiRoadExNet are higher than these of RoadExNet.

4.4.3. CHN6-CUG dataset

Table 4 reports the quantitative evaluation results of different methods in different labeled rate. Overall, SemiRoadExNet achieves the best performance again. SemiRoadExNet gets the highest F1, OA, and Kappa values except F1 values of 20% labeled rate. Meanwhile, RoadExNet gets better results than D-LinkNet and SII-Net. SemiRoadExNet gets higher 3.01% IoU, 3.35% Kappa and less 0.53% F1 than SII-Net. The improvements are 5.02%, 8.06% and 9.61% respectively when the labeled rate is 10% when labeled rate decreases to 5%, the improvement are 5.86%, 6.69% and 8.44% respectively. Compared with FMWDCT, SemiRoadExNet gets higher improvements with lower labeled rate.

As for comparison between baselines, D-LinkNet performances better than SII-Net except 20% labeled rate. SII-Net gets 5.41%, 2.88% and 3.28% improvements of F1, IoU and Kappa respectively, when labeled rate is 20%. Note that, s4GAN achieves worse performance than SII-Net. This is because SII-Net contains SIIS module for road extraction, and MLMT branch of s4GAN is effective in binary semantic segmentation. When the labeled rate is 10%, SII-Net gets 10.16%, 9.99% and 11.08% improvements respectively compared to s4GAN. Compared to other baselines, performances of ST++ are the worst in F1, IoU and Kappa metrics. This is because ST++ relies high-quality pseudo labels much. However, when the number of labeled data is limited, pseudo labels are not reliable enough for the training process of ST++. Also, the roads in this dataset are hard to distinguish due to their complex situation.

The results of different methods with 20% labeled rate are shown in Fig. 11. The road area are wide compared to image width and mainly lies in city area, which means complex road conditions. Many missed road areas and false road areas are in the comparative methods, and proposed SemiRoadExNet achieves the best visual performance. Note that part of the road labels is different from preliminary judgment of human eyes in the second row and the last row. This confirms that the roads in the dataset are hard to be distinguished due to complex road conditions. Also, maybe more labeling errors are in the dataset compared to Deep Globe Road Extraction dataset Massachusetts Roads dataset. In general, the road predictions of SemiRoadExNet are closer to ground truth images.

4.5. Ablation study

In order to validate the effects of D_s, D_e, attention module and rotation consistency constraint loss l_{rcc}, we make some ablation study experiments. We use the three road extraction datasets, and the labeled rate is set as 10%.

From Table 5, we can see the ablation study results. The IoU decreases 3.70%, 1.44% and 1.24% without D_s in DeepGlobe Road Extraction dataset, Massachusetts Roads dataset and CHN6-CUG dataset respectively. The result confirms D_s is beneficial to improve the performance of the generator. D_s makes the generator improve road

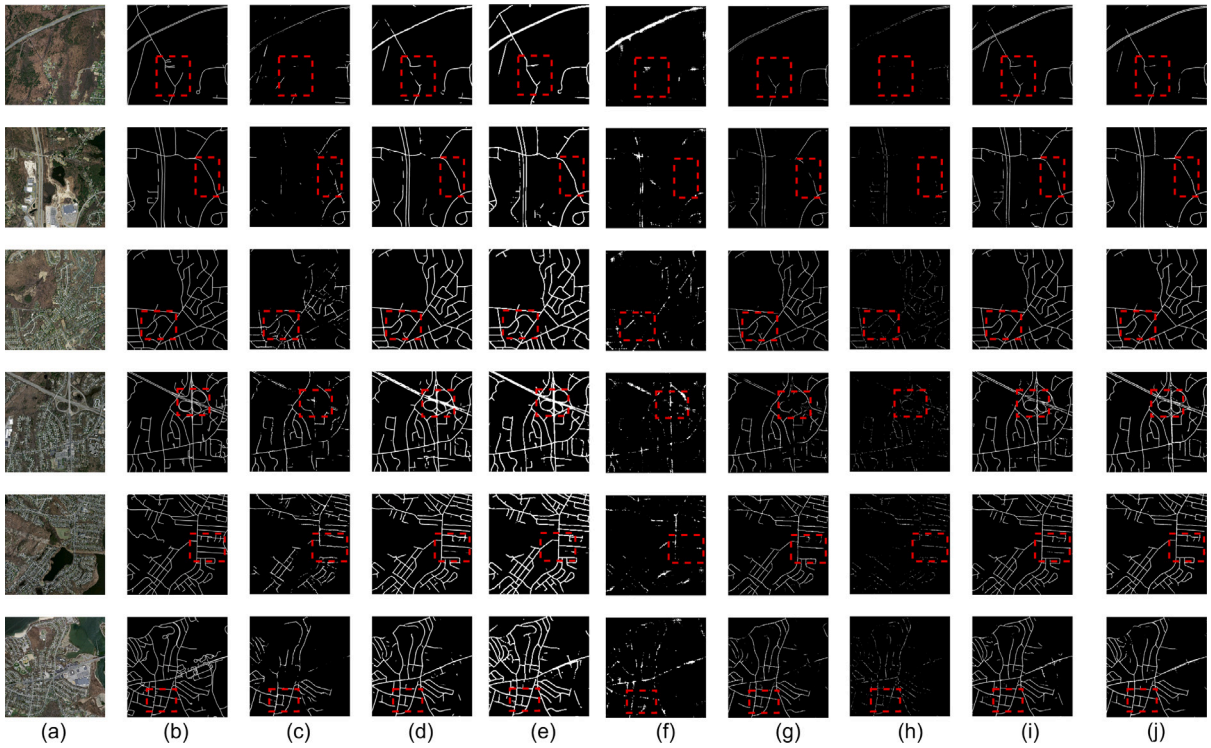

Fig. 10. Visual comparisons of different methods on Massachusetts Roads dataset with 20% labeled rate. (a) Remote sensing image. (b) Ground truth road image. (c) D-LinkNet. (d) SII-Net. (e) RoadExNet. (f) AdvNet. (g) s4GAN. (h) ST++. (i) FMWDCT. (j) SemiRoadExNet.

Table 4
Summary of the quantitative results on CHN6-CUG dataset.

Method	5%			10%			20%		
	F1	IoU	Kappa	F1	IoU	Kappa	F1	IoU	Kappa
D-LinkNet Zhou et al. (2018)	0.5648	0.3854	0.4700	0.6010	0.4295	0.5203	0.6195	0.4278	0.5118
SII-Net (Tao et al., 2019)	0.5594	0.3675	0.4410	0.5783	0.3682	0.4435	**0.6736**	0.4566	0.5446
RoadExNet	0.5880	0.4138	0.5017	0.6259	0.4383	0.5261	0.6435	0.4537	0.5422
AdvNet (Hung et al., 2019)	0.2695	0.1401	0.1601	0.2361	0.1228	0.1660	0.3252	0.1695	0.2214
s4GAN (Mittal et al., 2019)	0.3946	0.2092	0.2603	0.4767	0.2682	0.3326	0.4994	0.2704	0.3336
ST++(Yang et al., 2022)	0.0685	0.0054	0.0067	0.0893	0.0045	0.0048	0.0458	0.0012	0.0017
FMWDCT (You et al., 2022)	0.5600	0.3806	0.4717	0.6180	0.4295	0.5219	0.6629	0.4771	0.5705
SemiRoadExNet	**0.6180**	**0.4344**	**0.5253**	**0.6285**	**0.4488**	**0.5396**	0.6683	**0.4867**	**0.5781**

Table 5
Results for ablation study on three datasets.

Method	DeepGlobe Road Extraction			Massachusetts Roads			CHN6-CUG		
	F1	IoU	Kappa	F1	IoU	Kappa	F1	IoU	Kappa
Without D_s	0.6332	0.4775	0.6140	0.6886	0.5301	0.6724	0.6277	0.4363	0.5277
Without D_e	0.6337	0.4792	0.6161	0.6933	0.5358	0.6786	0.6070	0.4192	0.5075
Without attention module	0.5555	0.3962	0.5303	0.5355	0.3663	0.5035	0.5951	0.4058	0.4908
Without l_{rcc}	0.5709	0.4093	0.5439	0.5423	0.3727	0.5105	0.6256	0.4412	0.5307
SemiRoadExNet	**0.6674**	**0.5145**	**0.6304**	**0.6995**	**0.5445**	**0.6859**	**0.6285**	**0.4488**	**0.5396**

segmentation ability by comparing unlabeled image prediction images and labeled images prediction images. The IoU decreases 3.53%, 0.90% and 2.96% without D_e in DeepGlobe Road Extraction dataset, Massachusetts Roads dataset and CHN6-CUG dataset respectively. The results confirm that D_e allows the proposed method to utilize potential information of low-confidence pseudo labels from unlabeled data. D_e restricts the entropy distribution, therefore promoting the feature distribution consistency between unlabeled and labeled entropy maps. Therefore, the generator combines with D_s and D_e performs best. The IoU decreases 11.83%, 17.83% and 4.29% without attention module in DeepGlobe Road Extraction dataset, Massachusetts Roads dataset and CHN6-CUG dataset respectively. The performance drops significantly, which reflects the importance of the attention module for our model. The channel attention assigns weights to different channels so that makes road features more distinguishable. The non-local special attention utilizes the linear characteristic of road to mitigate label imbalance problem in some extent. Therefore, the attention module effectively improves road extraction performance. The IoU decreases 10.52%, 17.18% and 0.75% without l_{rcc} in DeepGlobe Road Extraction dataset, Massachusetts Roads dataset and CHN6-CUG dataset respectively. Note that roads in Massachusetts Roads dataset are thinnest and roads in CHN6-CUG dataset are widest, the rotation consistency constraint makes network predictions more accurate especially when the roads are very thin.

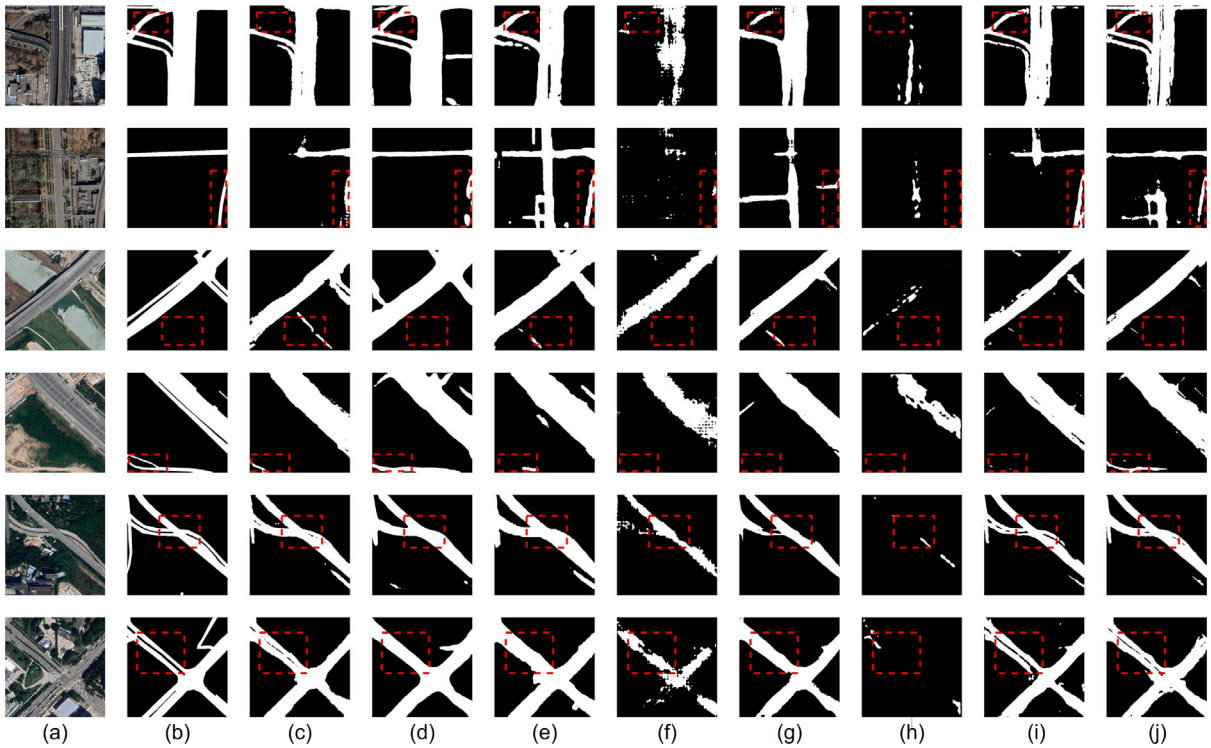

Fig. 11. Visual comparisons of different methods on CHN6-CUG dataset with 20% labeled rate. (a) Remote sensing image. (b) Ground truth road image. (c) D-LinkNet. (d) SII-Net. (e) RoadExNet. (f) AdvNet. (g) s4GAN. (h) ST++. (i) FMWDCT. (j) SemiRoadExNet.

5. Discussion

Compared to typical architecture of GAN, we design an additional discriminator branch D_e in our road extraction model to make a full use of the potential information of high-uncertainty pixels in pseudo labels. In order to visually show the function of D_e branch, we draw entropy maps and confidence maps for some road segmentation images, which are shown in Fig. 12. In entropy maps (Fig. 12(c) and (d)), the low-confidence area is red zone. The low-confidence area is small and exists in the edge of road (non-road) area. From Fig. 12(c) and (d), we can see there are more confidence area in the entropy map with D_e branch. Therefore, the predicted road edge area is more certain with D_e branch. We also can see, with D_e branch, the entropy map is more similar to ground truth road image, which means the road extraction ability is higher. From Fig. 12(a), we can see some trees exist on the road. Naturally, there are some truncated areas in road prediction images. Form Fig. 12(c) and (d) we can see, the truncated areas are larger without D_e, while there is hardly no truncation with D_e branch. In Fig. 12(e) and (f), low confidence area is marked as green. And we can see the low confidence area is generally smaller with D_e. Therefore, it reflects that D_e branch can overcome the influence of obstacles on distinguishment in some extent. From Fig. 12, we can conclude that D_e branch can reduce low-confidence area of road, and it can lift network performance.

To further illustrate the effect of D_e, Fig. 13 shows t-SNE 2-D embedding feature distribution of the network with and without D_e. The features are generated before classification layer (i.e., the last convolution layer). Compare to the feature distribution with D_e, there are more road feature dots mixed in road feature clusters without D_e. Therefore, from Fig. 13, we can see D_e makes embedding feature distribution of road and non-road more discriminating.

Shown As Figs. 12, 13 and Table 5, by adding an entropy discriminator branch can improve the performance of GAN based model. Since the entropy discriminator branch has no additional requirements for GAN network structure, similar entropy discriminator structure can be designed and added to other remote sensing imagery analysis and interpretation tasks such as building boundary extraction, water body detection, land use classification, and so on.

However, for different applications, the entropy discriminator branch needs to be designed delicately. Moreover, with the extra entropy discriminator branch, GAN model needs more calculation resource and memory. The risk of mode collapse which is very common in the training process of GAN models increases, since the layers of the network are deeper and the architecture of the model is more complex.

6. Conclusion

Recent supervised road extraction based on deep learning relies on much well annotated data. However, the labeled data is hard to acquire because annotation for road extraction task is time-consuming and costlier. Current semi-supervised road extraction models always underutilize the latent information of low-confidence pixels of pseudo-labels. To address such an issue, in this article we propose a novel semi-supervised road extraction network named SemiRoadExNet based on GAN. SemiRoadExNet contains a generator and two discriminators. The generator is an encoder–decoder structure with attention module. The two discriminators are both based on UNet. The generator generates road segmentation images and the according entropy maps. One discriminator inputs road predictions to encourage segmentation output feature distribution consistency. The other discriminator inputs prediction entropy maps to suppresses low-confidence areas for the unlabeled images, which mines entropy distribution information. Through GAN training, benefits of the pseudo labels and rotation consistency constraint, the generator gains information from labeled data and unlabeled data. The effectiveness and reliability of the proposed method have been verified on three different datasets. The experimental results show that the proposed SemiRoadExNet gains competitive performances compared to the state-of-the-art methods.

In future research, we will pay attention to designing more effective attention mechanisms and adopting other semi-supervised learning

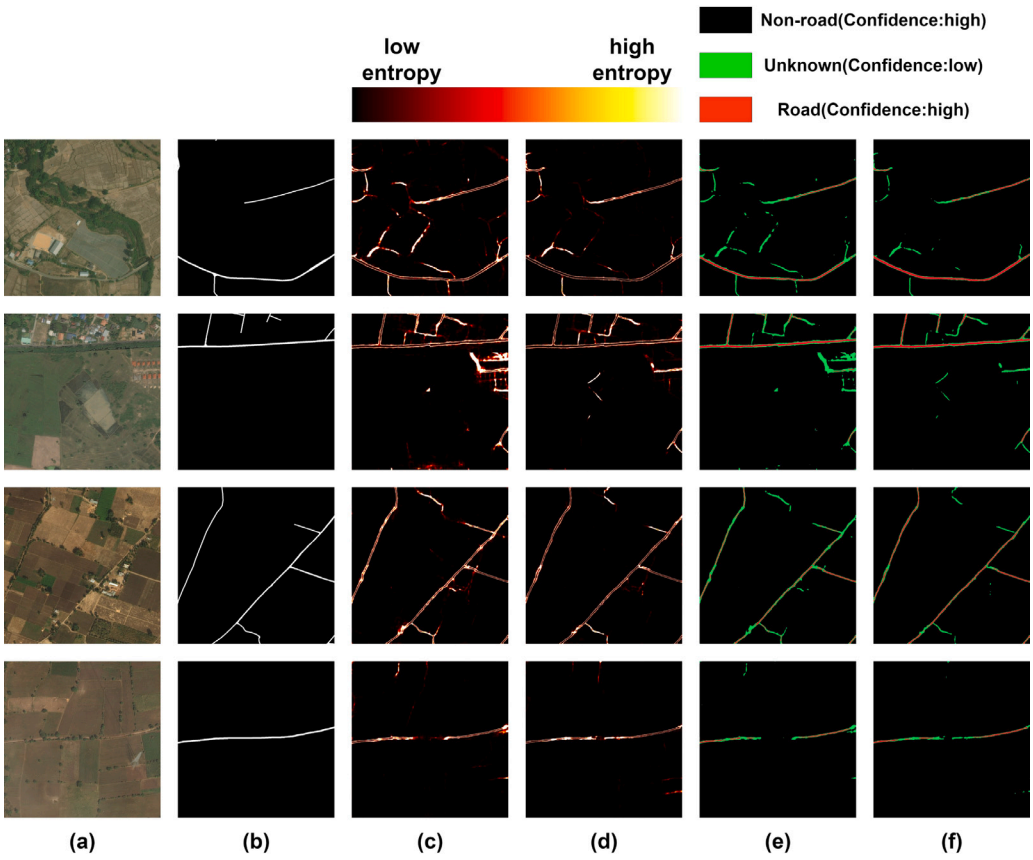

Fig. 12. The entropy maps of results. (a) Remote sensing image. (b) Ground truth road image. (c) The entropy map without D_e. (d) The entropy map with D_e. (e) High confidence area and low confidence area without D_e. (f) High confidence area and low confidence area with D_e. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

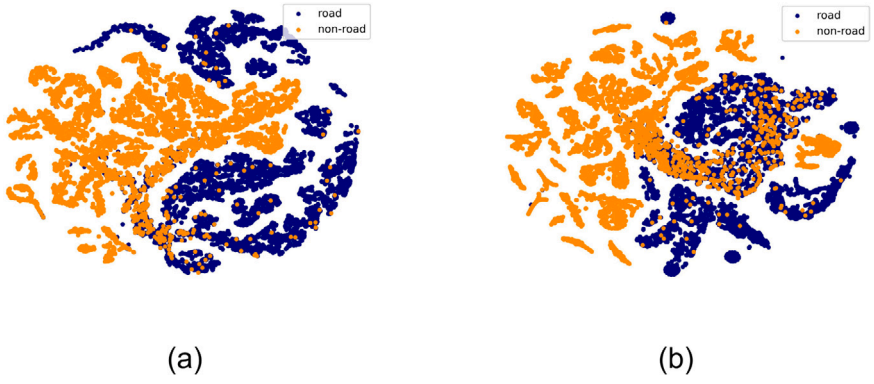

Fig. 13. Visualization of embedding features after applying t-SNE on DeepGlobe Road Extraction dataset with 10% labeled rate. The features are all from the feature before classification layer. (a) Embedding feature distribution with D_e. (b) Embedding feature distribution without D_e.

methods such as mean teacher and graph neural networks. We also would like to focus on road extraction using 3-D data, in order to get more practical road networks for our daily life.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. This work is supported in part by the National NSF of China under grants No. U19A2058, No. 41971362, No. 41871248 and No. 62106276.

References

Abdollahi, A., Bakhtiari, H.R.R., Nejad, M.P., 2018. Investigation of SVM and level set interactive methods for road extraction from google earth images. J. Indian Soc. Remote Sens. 46 (3), 423–430.

Abdollahi, A., Pradhan, B., Sharma, G., Maulud, K.N.A., Alamri, A., 2021. Improving road semantic segmentation using generative adversarial network. IEEE Access 9, 64381–64392.

Abdollahi, A., Pradhan, B., Shukla, N., 2019. Extraction of road features from UAV images using a novel level set segmentation approach. Int. J. Urban Sci. 23 (3), 391–405.

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., Alamri, A., 2020. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. Remote Sens. 12 (9), 1444.

Alshehhi, R., Marpu, P.R., 2017. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. ISPRS J. Photogramm. Remote Sens. 126, 245–260.

Bong, D.B., Lai, K.C., Joseph, A., 2009. Automatic road network recognition and extraction for urban planning. Int. J. Appl. Sci. Eng. Technol. 5 (1), 209–215.

Chang, C.-K., Siagian, C., Itti, L., 2012. Mobile robot vision navigation based on road segmentation and boundary extraction algorithms. J. Vis. 12 (9), 200.

Chen, Z., Deng, L., Luo, Y., Li, D., Junior, J.M., Gonçalves, W.N., Nurunnabi, A.A.M., Li, J., Wang, C., Li, D., 2022a. Road extraction in remote sensing data: A survey. Int. J. Appl. Earth Obs. Geoinf. 112, 102833.

Chen, H., Peng, S., Du, C., Li, J., Wu, S., 2022b. SW-GAN: Road extraction from remote sensing imagery using semi-weakly supervised adversarial learning. Remote Sens. 14 (17), 4145.

Chen, Z., Wang, C., Li, J., Xie, N., Han, Y., Du, J., 2021a. Reconstruction bias U-Net for road extraction from optical remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2284–2294.

Chen, X., Yuan, Y., Zeng, G., Wang, J., 2021b. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622.

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. Proc. IEEE 104 (11), 2207–2219.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223.

Dai, Z., Yang, Z., Yang, F., Cohen, W.W., Salakhutdinov, R., 2017. Good semi-supervised learning that requires a bad GAN. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6513–6523.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 172–181.

Desai, S., Ghose, D., 2022. Active learning for improved semi-supervised semantic segmentation in satellite images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 553–563.

He, C., Liao, Z.-x., Yang, F., Deng, X.-p., Liao, M.-s., 2012. Road extraction from SAR imagery based on multiscale geometric analysis of detector responses. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (5), 1373–1382.

He, Y., Wang, J., Liao, C., Shan, B., Zhou, X., 2022. ClassHyPer: ClassMix-based hybrid perturbations for deep semi-supervised semantic segmentation of remote sensing imagery. Remote Sens. 14 (4), 879.

He, H., Yang, D., Wang, S., Wang, S., Li, Y., 2019. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. Remote Sens. 11 (9), 1015.

Hu, A., Chen, S., Wu, L., Xie, Z., Qiu, Q., Xu, Y., 2021. WSGAN: an improved generative adversarial network for remote sensing image road network extraction by weakly supervised processing. Remote Sens. 13 (13), 2506.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W., 2019. Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 603–612.

Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H., 2019. Adversarial learning for semi-supervised semantic segmentation. In: 29th British Machine Vision Conference. BMVC 2018.

Kirthika, A., Mookambiga, A., 2011. Automated road network extraction using artificial neural network. In: 2011 International Conference on Recent Trends in Information Technology. ICRTIT, IEEE, pp. 1061–1065.

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., et al., 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS J. Photogramm. Remote Sens. 115, 119–133.

Li, Y., Shi, T., Zhang, Y., Chen, W., Wang, Z., Li, H., 2021a. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. ISPRS J. Photogramm. Remote Sens. 175, 20–33.

Li, J., Sun, B., Li, S., Kang, X., 2021b. Semisupervised semantic segmentation of remote sensing images with consistency self-training. IEEE Trans. Geosci. Remote Sens. 60, 1–11.

Lian, R., Huang, L., 2021. Weakly supervised road segmentation in high-resolution remote sensing images using point annotations. IEEE Trans. Geosci. Remote Sens. 60, 1–13.

Lian, R., Wang, W., Mustafa, N., Huang, L., 2020. Road extraction methods in high-resolution remote sensing images: A comprehensive review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5489–5507.

Liu, P., Wang, Q., Yang, G., Li, L., Zhang, H., 2022. Survey of road extraction methods in remote sensing images based on deep learning. PFG–J. Photogramm. Remote Sens. Geoinformation Sci. 90 (2), 135–159.

Lu, X., Zhong, Y., Zheng Zhuo, D., Su, A., Zhang, L., 2022. Cascaded multi-task road extraction network for road surface, centerline, and edge extraction. IEEE Trans. Geosci. Remote Sens. 60, 1–14.

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., Jie, W., 2015. Remote sensing big data computing: Challenges and opportunities. Future Gener. Comput. Syst. 51, 47–60.

Manandhar, P., Marpu, P.R., Aung, Z., 2018. Deep learning approach to update road network using VGI data. In: 2018 International Conference on Signal Processing and Information Security. ICSPIS, IEEE, pp. 1–4.

Manandhar, P., Marpu, P.R., Aung, Z., Melgani, F., 2019. Towards automatic extraction and updating of VGI-based road networks using deep learning. Remote Sens. 11 (9), 1012.

Miao, Z., Shi, W., Zhang, H., Wang, X., 2012. Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines. IEEE Geosci. Remote Sens. Lett. 10 (3), 583–587.

Miao, Z., Wang, B., Shi, W., Zhang, H., 2014. A semi-automatic method for road centerline extraction from VHR images. IEEE Geosci. Remote Sens. Lett. 11 (11), 1856–1860.

Mittal, S., Tatarchenko, M., Brox, T., 2019. Semi-supervised semantic segmentation with high-and low-level consistency. IEEE Trans. Pattern Anal. Mach. Intell. 43 (4), 1369–1379.

Miyamoto, R., Nakamura, Y., Adachi, M., Nakajima, T., Ishida, H., Kojima, K., Aoki, R., Oki, T., Kobayashi, S., 2019. Vision-based road-following using results of semantic segmentation for autonomous navigation. In: 2019 IEEE 9th International Conference on Consumer Electronics. ICCE-Berlin, IEEE, pp. 174–179.

Mnih, V., Hinton, G.E., 2010. Learning to detect roads in high-resolution aerial images. In: European Conference on Computer Vision. Springer, pp. 210–223.

Peddinti, A.S., Chouhan, A.S., Panigrahy, A.K., 2021. Road extraction using aerial images for future navigation. Mater. Today Proc. 47, 6306–6308.

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2020. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Trans. Geosci. Remote Sens. 59 (7), 5891–5906.

Sghaier, M.O., Lepage, R., 2015. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (5), 1946–1958.

Shamsolmoali, P., Zareapoor, M., Zhou, H., Wang, R., Yang, J., 2020. Road segmentation for remote sensing images using adversarial spatial pyramid networks. IEEE Trans. Geosci. Remote Sens. 59 (6), 4673–4688.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.-L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Adv. Neural Inf. Process. Syst. 33, 596–608.

Song, M., Civco, D., 2004. Road extraction using SVM and image segmentation. Photogramm. Eng. Remote Sens. 70 (12), 1365–1371.

Song, J., Li, J., Chen, H., Wu, J., 2021. MapGen-GAN: a fast translator for remote sensing image to map via unsupervised adversarial learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2341–2357.

Song, J., Li, J., Chen, H., Wu, J., 2022. RSMT: A remote sensing image-to-map translation model via adversarial deep transfer learning. Remote Sens. 14 (4), 919.

Sun, C., Wu, J., Chen, H., Du, C., 2022. SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using siamese networks with graph attention. Remote Sens. 14 (12), 2801.

Tao, C., Qi, J., Li, Y., Wang, H., Li, H., 2019. Spatial information inference net: Road extraction using road-specific contextual information. ISPRS J. Photogramm. Remote Sens. 158, 155–166.

Tao, Y., Xu, M., Zhang, F., Du, B., Zhang, L., 2017. Unsupervised-restricted deconvolutional neural network for very high resolution remote-sensing image classification. IEEE Trans. Geosci. Remote Sens. 55 (12), 6805–6823.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440.

Wang, S., Cao, J., Philip, S.Y., 2022a. Deep learning for spatio-temporal data mining: A survey. IEEE Trans. Knowl. Data Eng. 34 (08), 3681–3700.

Wang, J.-X., Chen, S.-B., Ding, C.H., Tang, J., Luo, B., 2021. RanPaste: Paste consistency and pseudo label for semisupervised remote sensing image semantic segmentation. IEEE Trans. Geosci. Remote Sens. 60, 1–16.

Wang, J., HQ Ding, C., Chen, S., He, C., Luo, B., 2020a. Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label. Remote Sens. 12 (21), 3603.

Wang, M., Luo, C., 2005. Extracting roads based on Gauss Markov random field texture model and support vector machine from high-resolution RS image. IEEE Trans. Geosci. Remote Sens. 9, 271–276.

Wang, Y., Peng, Y., Li, W., Alexandropoulos, G.C., Yu, J., Ge, D., Xiang, W., 2022b. Ddu-net: Dual-decoder-u-net for road extraction using high-resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–12.

Wang, J., Qin, Q., Yang, X., Wang, J., Ye, X., Qin, X., 2014. Automated road extraction from multi-resolution images using spectral information and texture. In: 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, pp. 533–536.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020b. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Seattle, WA, USA, pp. 13–19.

Wei, Y., Ji, S., 2021. Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–12.
Wei, Y., Wang, Z., Xu, M., 2017. Road structure refined CNN for road extraction in aerial image. IEEE Geosci. Remote Sens. Lett. 14 (5), 709–713.
Wu, S., Du, C., Chen, H., Xu, Y., Guo, N., Jing, N., 2019. Road extraction from very high resolution images using weakly labeled OpenStreetMap centerline. ISPRS Int. J. Geo-Inf. 8 (11), 478.
Xin, J., Zhang, X., Zhang, Z., Fang, W., 2019. Road extraction of high-resolution remote sensing images derived from DenseUNet. Remote Sens. 11 (21), 2499.
Xu, Y., Chen, H., Du, C., Li, J., 2021. Msacon: Mining spatial attention-based contextual information for road extraction. IEEE Trans. Geosci. Remote Sens. 60, 1–17.
Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y., 2022. St++: Make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4268–4277.
You, Z.-H., Wang, J.-X., Chen, S.-B., Tang, J., Luo, B., 2022. FMWDCT: Foreground mixup into weighted dual-network cross training for semisupervised remote sensing road extraction. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 5570–5579.
Yue, K., Yang, L., Li, R., Hu, W., Zhang, F., Li, W., 2019. TreeUNet: Adaptive tree convolutional neural networks for subdecimeter aerial image segmentation. ISPRS J. Photogramm. Remote Sens. 156, 1–13.
Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.
Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., Shinozaki, T., 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. Adv. Neural Inf. Process. Syst. 34, 18408–18419.
Zhang, B., Zhang, Y., Li, Y., Wan, Y., Wen, F., 2020. Semi-supervised semantic segmentation network via learning consistency for remote sensing land-cover classification. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2, 609–615.
Zheng, Y., Yang, M., Wang, M., Qian, X., Yang, R., Zhang, X., Dong, W., 2022. Semi-supervised adversarial semantic segmentation network using transformer and multiscale convolution for high-resolution remote sensing imagery. Remote Sens. 14 (8), 1786.
Zhou, J., Bischof, W.F., Caelli, T., 2006. Road tracking in aerial images based on human–computer interaction and bayesian filtering. ISPRS J. Photogramm. Remote Sens. 61 (2), 108–124.
Zhou, L., Zhang, C., Wu, M., 2018. 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. pp. 182–186.
Zhu, Q., Zhang, Y., Wang, L., Zhong, Y., Guan, Q., Lu, X., Zhang, L., Li, D., 2021. A global context-aware and batch-independent network for road extraction from VHR satellite imagery. ISPRS J. Photogramm. Remote Sens. 175, 353–365.
Zou, Y., Zhang, Z., Zhang, H., Li, C.-L., Bian, X., Huang, J.-B., Pfister, T., 2020. Pseudoseg: Designing pseudo labels for semantic segmentation. arXiv preprint arXiv:2010.09713.