This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3053603, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Reconstruction Bias U-Net for Road Extraction from Optical Remote Sensing Images

Ziyi Chen, *Member, IEEE*, Cheng Wang, *Senior Member, IEEE* , Jonathan Li, *Senior Member, IEEE*, Nianci Xie, Yan Han and Jixiang Du*, *Member, IEEE*

***Abstract*—Automatic road extraction from remote sensing images plays an important role for navigation, intelligent transportation and road network update etc. Convolutional Neural Network (CNN) based methods have presented many achievements for road extraction from remote sensing images. CNN based methods require a large dataset with high quality labels for model training. However, there is still few standard and large dataset which is specially designed for road extraction from optical remote sensing images. Besides, the existing end-to-end CNN models for road extraction from remote sensing images are usually with symmetric structure, studying on asymmetric structure between encoding and decoding is rare. To address the above problems, this paper first provides a publicly available dataset LRSNY for road extraction from optical remote sensing images with manually labelled labels. Second, we propose a reconstruction bias U-Net for road extraction from remote sensing images. In our model, we increase the decoding branches to obtain multiple semantic information from different up-samplings. Experimental results show that our method achieves better performance compared with other six state-of-the-art segmentation models when testing on our LRSNY dataset. We also test on Massachusetts and Shaoshan datasets. The good performances on the two datasets further prove the effectiveness of our method.**

***Index Terms*—Road extraction, optical remote sensing image, CNN model, dataset**

I. Introduction

BENEFITING from the prosperity and development of navigation, automatic driving, smart city and intelligent transportation etc., road network information plays a more and more important role in our daily life. Due to the new road construction, road network information update is always necessary. There are many kinds of methods for road network information update, such as manually labeling, tracking the changes of cars' driving traces and automatic road extraction from remote sensing images etc. Tracking cars' driving traces may miss the appearance information of roads, such as width, border and so on. Manually labeling is time consuming and with hard manual burden. Automatic road extraction from optical remote sensing images is a more economic and more time saving way compared with the traditional manual road areas labeling [1]. Precisely because the attracting of high research values about road extraction from remote sensing images, much research has focused on automatic road extraction from remote sensing images [1-8].

Most state-of-the-art road extraction methods from remote sensing images are CNN models based. Especially, due to the fast development of end-to-end semantic segmentation CNN models in natural images, the end-to-end semantic segmentation CNN models also achieve great success in road extraction from remote sensing images. However, there are still two aspects that need to be considered when using CNN segmentation models for road extraction from remote sensing images. First, CNN models need a large dataset with labels for training. For now, there is still few publicly available dataset which is large enough, well labeled and specially designed for road extraction from remote sensing images. Constructing a large dataset for road extraction from remote sensing images with manually labelled labels is costly. Second, most current end-to-end CNN segmentation models use symmetric or approximate symmetric structures between encoding part and decoding part, such as U-Net[9], PSPNet[10] etc. However, as reconstruction is a more challenging job compared with feature extraction work, using neural network parts with similar complexity to finish jobs with different difficulties seems unreasonable, which may result to the unbalance between reconstruction capacity and feature extraction capacity.

To overcome the above two problems of road extraction from remote sensing images, this paper's work major focuses on two aspects. First, we cost a major expenditure of time and effort to label a large dataset for road extraction from remote sensing images. The dataset is not only well labeled, but also large enough to train CNN models and obtain sound test evaluations. To make experiments which using our dataset for training and testing be fair among different CNN models, we have divided our dataset into constant training, validation and testing parts. Second, we propose a reconstruction bias U-Net for road extraction from optical remote sensing images. In our reconstruction bias U-Net, we use multiple operation couples of upsampling and convolution to increase the reconstruction ability for each upsampling layer. Finally, through increasing reconstruction operations, we can strengthen the reconstruction ability of CNN model and achieve better segmentation results in the experiments.

The contributions of this paper lie on:

Manuscript received November 12, 2020.

Z. Chen, Nianci Xie, Yan Han and J. Du are with the Department of Computer Science and Technology, Huaqiao University, Xiamen, JS 361021, China (e-mail: chenziyihq@hqu.edu.cn; 592546548@qq.com; 1010786675@qq.com; and jxdu@hqu.edu.cn).

C. Wang is with the School of Information Science and Engineering, Xiamen University, Xiamen, FJ 361005, China (e-mail: cwang@xmu.edu.cn).

J. Li is with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@uwaterloo.ca).

(1) We provide a publicly available large dataset with manually labelled labels for road extraction from remote sensing images.
(2) We propose a reconstruction bias U-Net for road extraction from remote sensing images. Benefiting from increasing reconstruction ability, our new CNN model performs better compared with other state-of-the-art semantic segmentation CNN models.

II. Related Work

In this section, we give a review about datasets of semantic segmentation at first. In the second part, we give a detail review about road extraction from remote sensing images.

A. Studies on Datasets for Semantic Segmentation

Since most current state-of-the-art semantic segmentation methods are CNN models based, a large enough dataset with manually labelled labels is rather important for specific target segmentation researches. Many researchers have pay their large attention on and taken large efforts for semantic segmentation dataset construction [11-24]. Brostwo et al. proposed first semantic segmentation video dataset called Cambridge-driving Labeled Video Database (CamVid) [19]. In this dataset, 32 classes were labeled in the videos with 30 Hz for more than 10 minutes. Cordts et al. proposed Cityscapes dataset, which includes 5000 street scenes and corresponding exquisite labels from 50 different cities [18]. Usually, the Cityscapes dataset can be split into 2975 training images, 500 validation images and 1525 test images for total 19 classes (including roads in street viewpoint). Besides, 20000 street scenes with coarse ground truth are also provided. Ros et al. proposed SYNTHIA dataset which is a large scale scene of virtual city, providing pixel-level labeling for 11 classes (including road areas) [20]. As the dataset was produced based on a virtual city scene, the images varies with different viewpoints, seasons and weather. Huang et al. proposed ApolloScape for automatic driving research. In ApolloScape dataset, more than 147 thousands frames with pixel-level semantic labeling are publicly open, coving 3 different cities. The above representative datasets are based on ground or street viewpoints, there are also many semantic segmentation dataset for remote sensing images. ISPRS dataset [17] provides 38 large aerial image patches with same size and their corresponding DSM data. The dataset greatly promote the CNN model research on DSM segmentation of remote sensing images. Demir et al. raised a challenge competition about semantic segmentation from remote sensing images, including road or street net extraction [15]. Since the competition was over, the download of the dataset is not available. Nigam et al provided a semantic segmentation dataset obtained by aerial planes flying at height ranging from 5 meters to 50 meters [14]. The dataset contains 3268 images with 11 labeled classes. Mohajerani et al. provided a dataset for cloud segmentation study, which including 38 Landsat images [12]. Each image contains 4 bands' information: red, green, blue and NIR. Schmitt et al. proposed a dataset for multi-spectral image fusion study based on deep learning [11]. Mnih provided an aerial image dataset for road extraction from remote sensing images [25]. Bastani et al. provided a road extraction dataset which using aerial images covering 24 sq km around 15 cities[26]. However, in their provided link, they told readers their dataset can't be publicly released due to copyrights. Cheng et al. told readers they will publicly open their road centerline extraction dataset[2]. We do not find the dataset download link in the paper. Zang et al. used Shaoshan dataset for road extraction, however, their dataset is not public available [27].

From the above introduction, we can know that many researchers have spent large energy on dataset construction for semantic segmentation based on deep learning. However, a dataset specifically for the study of road extraction from remote sensing images based on deep learning models is still scarce and much-needed.

B. Road Extraction from Remote Sensing Images

Road extraction from remote sensing images usually contains two subtasks: road area extraction and road centerline extraction[28, 29]. Road area extraction methods produce pixel-level labeling of roads[1, 4, 29-38], while skeletons of roads are extracted for road centerline extraction[8, 27, 28, 39-46].

As roads have outstanding shape feature compared with other ground targets, the morphological features are utilized for road extraction [4, 28, 31, 47]. With the development of machine learning methods, such as support vector machine (SVM), many researchers used machine learning methods combining with artificial designed features for road extraction from remote sensing images and obtained many achievements [34, 36]. Poullis proposed a no threshold framework which called Tensor-Cuts, and applied the framework for pre-processing of road extraction from satellite images since the framework is particularly suitable for linear features extraction[36]. Movaghati et al. proposed a road extraction method from satellite images using particle filtering (PF) and extended kalman filtering(EKF)[48]. The PF is combined with EKF to find best continuation of the road after an obstacle or junction, which achieved satisfactory results. Leninisha et al. presented a semi-automatic framework based on geometric active deformable model for road network extraction from high spatial remote sensing images. Different road junctions shape types were extracted using water flow technique, and they achieved good results on test images[49]. Lv et al. proposed an adaptive multi-feature (which containing color, local entropy and HSC features) sparsity-based model for road area extraction, and they achieved good results in the experiments[33].

Recently, deep convolutional neural networks (CNN) have led a series of breakthroughs for computer vision tasks [50-59]. CNN models also have achieved many success in road extraction from remote sensing images[8, 38, 46, 60-64]. Alshehhi et al. proposed a single patch-based CNN for extraction of roads and buildings from high-resolution remote sensing data[31]. Experiments were conducted on two challenging datasets to demonstrate the performance of the proposed network architecture. Cheng et al. used a cascaded end-to-end CNN for automatic road detection and centerline extraction, which obtained the state-of-the-art results in the experiments[28]. Zhang et al. used a semantic segmentation neural network which combines the strengths of residual learning and U-Net[65] for road area extraction from remote sensing images[29]. They achieved better results compared with other state-of-the-arts approaches. Chen et al. proposed a road extraction approach from remote sensing images, which

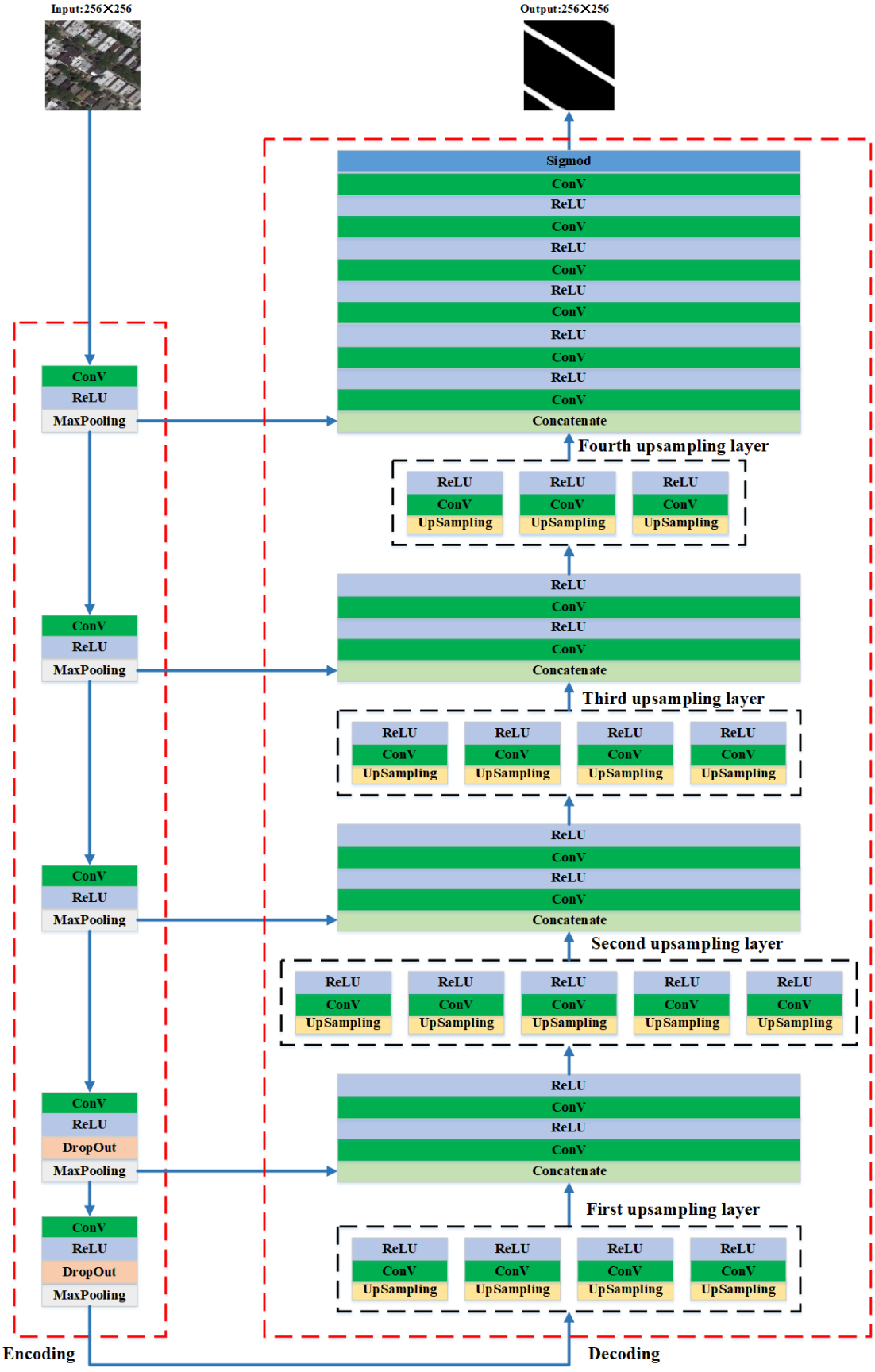


Fig. 1. The architecture of our reconstruction bias U-Net.

TABLE I. THE DETAIL NETWORK ARCHITECTURE OF RECONSTRUCTION BIAS U-NET.

	Operation couple	Filter	stride	Output size
Input				256×256×3
Encoding	[Conv, ReLu, MaxPooling] ×3	[3×3, ----, 2×2] × 3	[1, ----, 1] × 3	[128×128×64] [64×64×128] [32×32×256]
	[Conv, ReLu, Dropout, MaxPooling] ×2	[3×3, ----, 0.5, 3×3] ×2	[1, ----, 1] × 2	[16×16×512] [16×16×1024]
Decoding	[Up, Conv] + [Up, Conv] + [Up, Conv] + [Up, Conv]	[2×2, 2×2/512] + [2×2, 4×4/128] + [2×2, 8×8/64] + [2×2, 16×16/32]	[1, 1] × 4	[32×32×512] +[32×32×128] +[32×32×64] +[32×32×32]
	[Concat, Conv, ReLU, Conv, ReLu]	[----, 3×3/512, ----, 3×3/512, ----]	[----, 1, ----, 1, ----]	[32×32×512]
	[Up, Conv] + [Up, Conv] + [Up, Conv] + [Up, Conv] + [Up, Conv]	[2×2, 2×2/256] + [2×2, 4×4/64] + [2×2, 8×8/32] + [2×2, 16×16/16] + [2×2, 32×32/8]	[1, 1] × 5	[64×64×256] +[64×64×64] +[64×64×32] +[64×64×16] +[64×64×8]
	[Concat, Conv, ReLU, Conv, ReLu]	[----, 3×3/256, ----, 3×3/256, ----]	[----, 1, ----, 1, ----]	[64×64×256]
	[Up, Conv] + [Up, Conv] + [Up, Conv] + [Up, Conv]	[2×2, 2×2/128] + [2×2, 4×4/32] + [2×2, 8×8/16] + [2×2, 64×64/2]	[1, 1] × 4	[128×128×128] +[128×128×32] +[128×128×16] +[128×128×2]
	[Concat, Conv, ReLU, Conv, ReLu]	[----, 3×3/128, ----, 3×3/128, ----]	[----, 1, ----, 1, ----]	[128×128×128]
	[Up, Conv] + [Up, Conv] + [Up, Conv]	[2×2, 2×2/64] + [2×2, 4×4/16] + [2×2, 8×8/8]	[1, 1] × 3	[256×256×64] +[256×256×16] +[256×256×8]
	[Concat, Conv, ReLU, Conv, ReLu, Conv, ReLU, Conv, ReLU, Conv, ReLU, Sigmoid]	[----, 3×3/64, ----, 3×3/64, ----, 3×3/3, ----, 1×1/3, ----, 1×1/3, ----, 1×1/1]	[----, 1, ----, 1, ----, 1, ----, 1, ----, 1, ----, 1]	[256×256×1]

combines Dirichlet Mixture Models and CNN modes and achieved good results [1]. Ren et al. proposed a DA-CapsUNet for road extraction from remote sensing images [37]. In their approach, they used a capsule U-Net architecture to extract and fuse multiscale capsule features. They achieved quite good results in the experimental results.

III. METHOD

In this section, we show the CNN architecture of our reconstruction bias U-Net firstly. Then, we give a detail introduction about reconstruction bias part in our CNN model. Finally, we illustrate the loss training of our model.

A. Model Architecture of Reconstruction Bias U-Net

Fig. 1 shows the model architecture of our reconstruction bias U-Net. The major backbone of our network is based on the U-Net [9], consisting of encoding and decoding two parts. The initial input of our network is a 256×256 remote sensing image. In the encoding part, we use five groups of convolution, ReLU and maxpooling operations. In the fourth and fifth groups, dropout operation is added to reduce the activated neuron weights, avoiding the over fitting of the network. In the decoding part, we first use four operation couples of upsampling, convolution and ReLU to obtain four upsampling results. It should be emphasized that the kernel sizes of the four convolution operations are different, learning the reconstruction ability of using different sizes of context information. Then, we concatenate the four upsamping results of first upsampling layer with the output of max-pooling in the fourth operation couple of encoding part, followed by two operation couples of convolution and ReLU. In the second upsampling layer, we use five operation couples of upsampling, convolution and ReLU. In the next, the five upsampling results and the output of max-pooling in the third operation couple of encoding part are concatenated and followed by two operation couples of convolution and ReLU. The operations of the third upsampling layer is just same with the operations in the first upsampling layer. The outputs of third upsampling layer and the output of max-pooling in the second operation couple of encoding part are concatenated and followed by two operation couples of convolution and ReLU. In the fourth upsampling layer, three operation couples of upsampling, convolution and ReLU are used. The outputs of fourth upsampling layer and the max-pooling output in the first operation couple of encoding part are concatenated and followed by five operation couples of convolution and ReLU. In the final two operations, convolution and classification with sigmoid activation function are used. The output of final classification operation is the road area segmentation result with a size of 256×256.

B. Details of Decoding Architecture

In the reconstruction bias U-Net, we use multiple operation couples of upsampling and convolution to increase the reconstruction ability. Table I shows the detail information of inputs, outputs, filters and strides in each layer. From Table I we can see that through increasing the multiple operation couples of upsampling and convolution, our decoding part occupies much more parameters than encoding part. For each upsampling operation, the upsampling size is fixed at 2×2. There are 4, 5, 4 and 3 upsampling operations in the first, second, third and fourth upsampling layer, respectively. We design the four upsampling layers from the considering that middle layers need more upsampling operations to learn and joint more reconstruction information. The detail numbers are selected according to the ability of our GPU. If one has a more powerful GPU, he can add the upsampling numbers for each layer. In the first upsampling layer, the followed convolution operations have filter sizes of 2×2, 4×4, 8×8 and 16×16, respectively. The convolutional filter numbers are 512, 128, 64 and 32, respectively. The output size of first concatenation operation is 32×32×1248, followed by two convolutions with 512 filters having a 3×3 size. In the second upsampling layer, we use five couples of upsampling and convolution. The convolution operations have 256 2×2, 64 4×4, 32 8×8, 16 16×16, 8 32×32 filters, respectively. The output size of second concatenation operation is 64×64×632. In the third upsampling layer, we use four couples of upsampling and convolution. The convolution operations have 128 2×2, 32 4×4, 16 8×8, 2 64×64 filters, respectively. The output size of third concatenation operation is 128×128×306. In the final upsampling layer, we use three couples of upsampling and convolution. The convolution operations have 64 2×2, 16 4×4, 8 8×8 filters, respectively. The output size of fourth concatenation operation is 256×256×152. To further convert the output into 256×256×1, the output of fourth concatenation is followed by five couples of convolution and ReLU. The five convolution operations have 64 3×3, 64 3×3, 3 3×3, 3 1×1 and 3 1×1filters, respectively. In the final layer, we use a sigmoid to obtain the final segmentation result. It should be noted that, the numbers of upsampling and convolution operations in decoding part can be increased according to the GPU memory size.

C. Loss Function

Given a set of training images and the corresponding road area segmentation labels (I, G), the target function of the network can be represented as follow:

$$Min\ \mathrm{E}(I, G, W) = \sum_{i=1}^{N} ||I_i * W - G_i||^2, \qquad (1)$$

where N is the number of training images, W is the parameters of network. In our network training, we use binary cross entropy as the loss function:

$$\mathcal{L}_{Net_w}(I) = -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{L} G_i(j,k)\cdot logNet_w(I_i(j,k)) + (1 - G_i(j,k))\cdot \log(1 - Net_w(I_i(j,k))), \qquad (2)$$

where (M,L) represents the shape size of images I, $Net_w(I_i(j,k))$ represents the network output of position (j,k) in image I_i. We use the Adam (adaptive moment estimation) and accuracy metrics to train our network. Other loss functions (such as mean squared error, pixel-wise cross entropy etc.) and model training methods (such as stochastic gradient descent) are also can be used for training the network.

IV. RESULTS AND DISCUSSION

In this section, we will first give an introduction about the dataset used in experiments. Then, the detail experimental implementations are introduced. Third, we present and analyze the experimental results on the tested datasets.

A. Dataset

In this paper, we provide a publicly available road extraction dataset from high resolution remote sensing images. The original image is a 37949×35341 high resolution remote sensing image, covering a center part of New York City and with a resolution of 0.5m. We cut the original large image into pieces with a size of 1000×1000, generating 1368 images. The 1368 images are further divided into training, validation and test images. The training, validation and test images contain 716, 220 and 432 images, respectively.

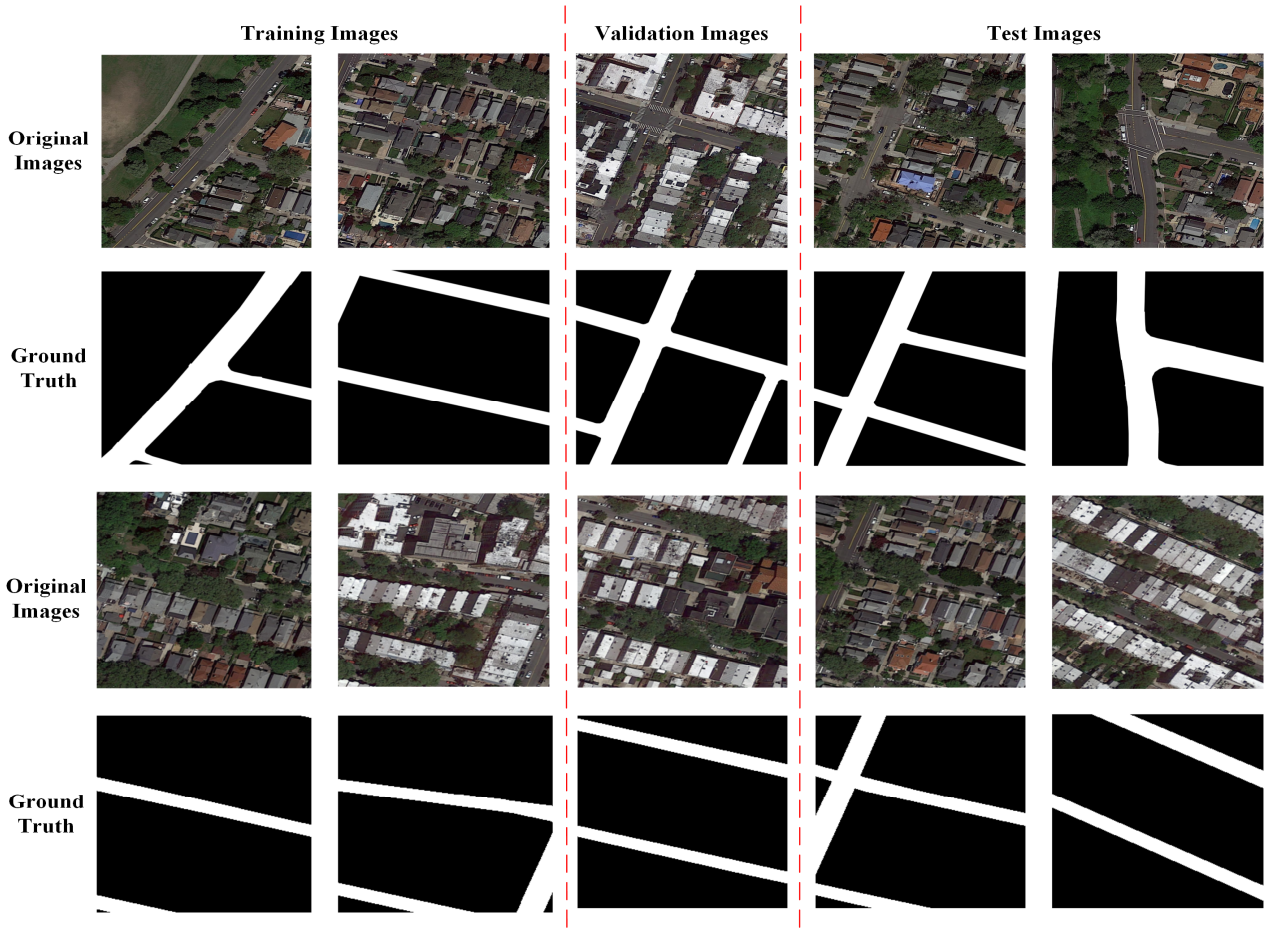


Fig. 2 Training, validation, test images and the corresponding ground truth exhibition of our dataset.

Compared with Massachusetts Roads Dataset[1], our dataset has a much higher resolution. The dataset size of our dataset is comparable with Massachusetts Roads Dataset. Massachusetts Roads Dataset contains 1108 training images, and each image has a size of 1500×1500. However, Massachusetts Roads Dataset only has 14 validation images and 49 test images. Compared with Cheng's road extraction dataset [2], our dataset is much larger. For Deepglobe road extraction dataset[2], the download was not available since the competition was over.

To maintain the position distribution balance of our dataset, we use a dividing strategy which iteratively assign training, validation and test images at column level. For example, given an image which are cut into 8×8 pieces, as shown in Fig. 3. Then, the pieces lay on first, fourth, sixth and eighth columns are divided into training set. The pieces lay on second, fifth and seventh columns are assigned to test set. The pieces lay on third column are assigned to validation set. Through the above dividing strategy, we can guarantee the position distribution balance of different kinds of image sets. It should be note that several validation images have small overlap areas with training images. The rest piece images at edge area of the original large image are usually not 1000 × 1000. To utilize the rest edge areas, we fill the pieces into 1000 × 1000 using small overlaps with nearby column pieces (in our dataset are training pieces). The test images have no overlaps with training and validation images.

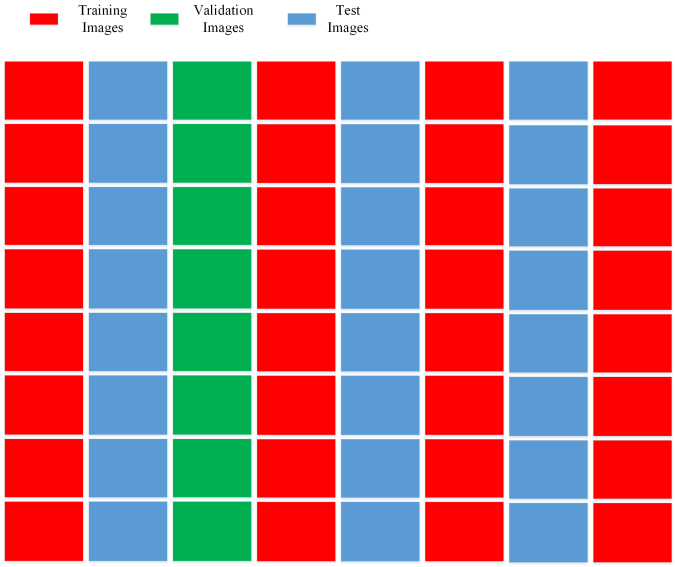


Fig. 3 The division strategy of our original large image. After division, the dataset is divided into training, validation and test datasets.

We manually label the ground truth of road areas in the original large image and implement the division as same as the

[1] http://www.cs.toronto.edu/~vmnih/data/

[2] https://competitions.codalab.org/competitions/18467

original image. Thus, it is one-to-one correspondence between the original image pieces and the ground truth label pieces. Because it will put a lot of pressure on the GPU memory size if we use training images with a size of 1000 × 1000, so we resize the images in our dataset into a size of 256 × 256. Although we only use training, validation and test images with a size of 256 × 256 in this paper, the images with a size of 1000 × 1000 are also reserved and publicly available in our dataset.

Fig. 2 shows several images and their corresponding ground truth in our dataset, including training, validation and test datasets. From Fig. 2 we can see that the images in our dataset are rather challenging, as many road areas are occlude by trees, buildings and cars etc. It should be note that we use white color (255, 255, 255) and black color (0, 0, 0) to respectively represent the road areas and background areas in our labeling images. Besides, because the training, validation and test images have been divided and are constant, our dataset can be easily used for performance comparison. To make our dataset be easy to remember, we call our dataset as LRSNY (Large Road Segmentation Dataset from Optical Remote Sensing Images of New York). Our dataset can be download from the following website: ftp://154.85.52.76/LRSNY/.

To further verify the performance of the proposed reconstruction bias U-Net, we also compared the reconstruction bias U-Net with other methods on other two datasets: Massachusetts road extraction dataset and Shaoshan dataset [1].

For Massachusetts dataset, we divide the original training, validation and test images into 256×256 without overlappings, generating 27700 training images, 350 validation images and 1225 test images, respectively.

Shaoshan dataset is a Pleiades optical image covering parts of Shaoshan with an image size of 11125×7918. We follow the processing in [1], using 29 and 20 images for training and test respectively. Each image has a size of 1589×1131. To suit the input size of 256×256, we further divide each image into 256×256. For training images, the division between generated neighbor 256×256 images has a overlap. We generate 256×256 training images with a skip step of 10 pixels in both row and column directions. Finally, we have 14580 training images and 456 test images.

B. Experimental Implementation Detail

We train our model on a computer with Intel® Core™ i9-9900X 3.5 GHz and 128 GB memories. The computer has two GPUs, which type is RTX 2080 Ti with 11 GB GPU memories. During training and test, we use only one GPU. When training our model, we set the training epoch as 200 and the learning rate as 0.0001. Our training batch size is 2. For model saving, we save the model with minimum loss within 200 epochs.

Our implementation is based on Python, Tensorflow[3] and Keras[4]. To further strengthen the training stage and avoid the over fitting problem of model training, we utilize the data augmentation for training images. The rotation, zoom, shift, shear and flip operations are all used in our training data augmentation. In our experiment, the rotation range is set at 0.9, the width shift range and height shift range are set at 0.1. And the shear range and zoom range are also set at 0.1.

C. Evaluation Criteria

In this part, we give a brief introduction about the performance evaluation criteria used in our experiment. To comprehensively evaluate the performance of models, we use four evaluation criteria which are widely used for evaluating road segmentation performance. The first three criteria are completeness, correctness and quality, the representations are as follows:

$$\begin{aligned} \text{completeness} &= \frac{TP}{TP+FN} \\ \text{correctness} &= \frac{TP}{TP+FP}, \\ \text{quality} &= \frac{TP}{TP+FN+FP} \end{aligned} \tag{3}$$

where TP, FN and FP denote true positive, false negative and false positive, respectively.

The fourth evaluation criterion is PRI (Probabilistic Rand Index), which can be computed as follow:

$$\text{PRI}(S_{seg}, S_{gt}) = \frac{1}{C_n^2}\sum_i \sum_{j(i\neq j)} [\psi(l_i = l_j \& l_i' = l_j') + \psi(l_i \neq l_j \& l_i' \neq l_j')], \tag{4}$$

where ψ is a discrimination function, l_i and l_j are the labels of S_{seg}, l_i' and l_j' are the labels of S_{gt}, c_n is the total pixel numbers of S_{seg}.

D. Experimental Results

In this section, we first present the comparison results among our model and other state-of-the-art segmentation methods on LRSNY dataset. The compared methods include the original U-Net[9], SegNet [66], PSPNet [10], Residual U-Net[67], DeepLabV3[68] and DANet [69]. For each method, we train the model 200 epochs, which is same with our model training setting. Besides, to make the comparison fair, the training images augmentation operation and augmentation parameter settings are also just the same with our method for all the other comparing methods.

Table II shows the comparison results among our method and other six state-of-the-art segmentation methods tested on our LRSNY dataset. From table II, we can see that our method obtains best performance according to the quality scores, which proves the effectiveness of our reconstruction bias U-Net. It can be seen that our method obtains about 0.4%, 1.4%, 0.4%, 4.2%, 2%, 1% higher quality scores than the original U-Net, SegNet, PSPNet-50, Residual U-Net, DeepLabV3, DANet, respectively.

Table II. THE COMPARISON RESULTS AMONG OUR METHOD AND OTHER SIX STATE-OF-THE-ART SEGMENTATION METHODS TESTED ON OUR LRSNY DATASET.

Method	Completeness	Correctness	Quality
U-Net[9]	**0.9398**	0.91599	0.86523
SegNet[66]	0.91233	0.93219	0.85555
PSPNet-50[10]	0.91221	0.94351	0.86497
Residual U-Net[67]	0.90218	0.90899	0.82744
DeepLabV3[68]	0.90588	0.9323	0.84996
DANet[69]	0.90504	**0.94521**	0.85993
Ours	0.92143	0.93864	**0.86908**

[3] https://www.tensorflow.org/

[4] https://keras.io/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3053605, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

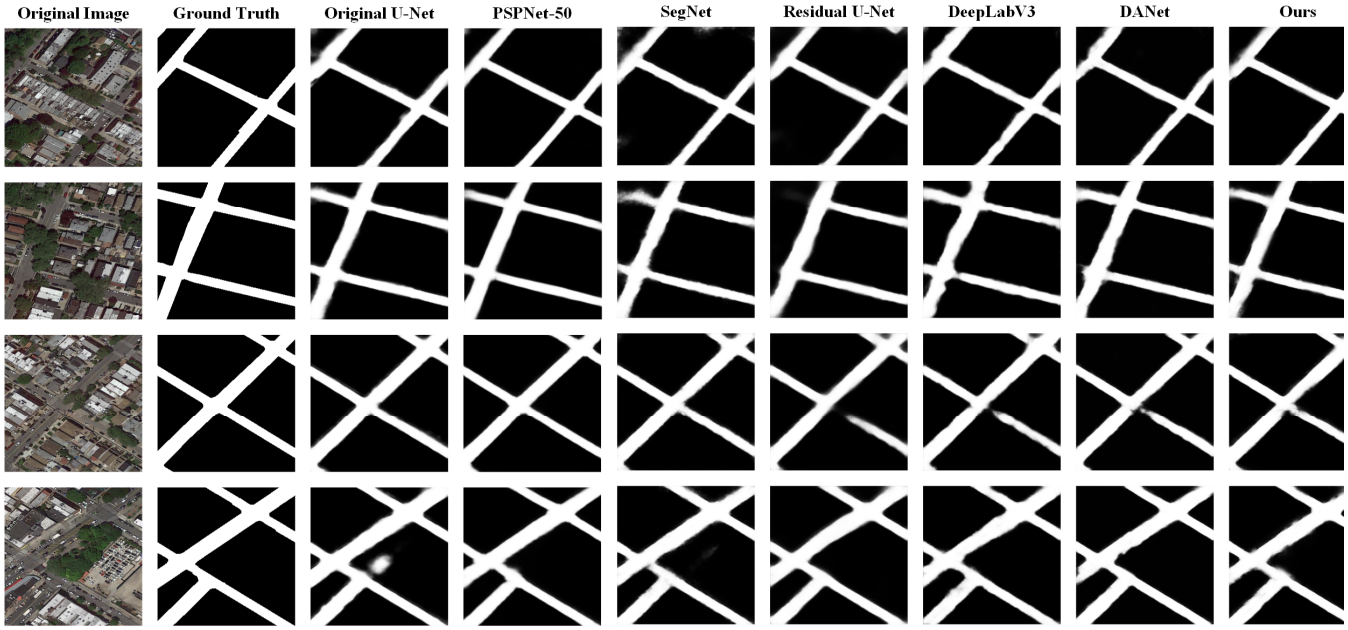


Fig. 4. Visiual segmentation results' exhibition of different methods tested on our LRSNY dataset.

Fig. 5 shows the PRI comparison among our method and other 6 state-of-the-art methods tested on LRSNY dataset. From Fig. 5, we can clearly see that our method obtains highest PRI score among all the methods, which further proves the good performance of our method.

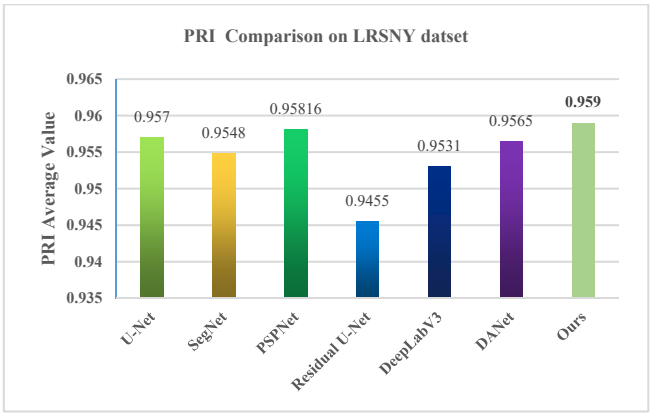


Fig. 5. The PRI comparison among our method and other six state-of-the-art methods tested on LRSNY dataset.

Fig. 4 shows the visual segmentation results of different methods tested on our LRSNY dataset. The first to eighth columns are the original images, ground truth, results of original U-Net, SegNet, Residual U-Net, DeepLabV3, DANet and ours, respectively. From the figure, we can see that the segmentation results of our method obtains a better balance and stable performance when dealing with different images. The reason is due to our reconstruction bias part in our model, resulting to better performance.

In the next part of this section, we present the comparison results among our reconstruction bias U-Net with other six methods on Massachusetts road extraction dataset. In the experiment, the compared methods are as same as the six methods compared in the prior part. The training implementation parameters are also same as the parameters in the prior experiment, except that we set training epoch as 20, steps per epoch as 27700 and validation steps as 350.

Table III shows the comparison results among our method and other six state-of-the-art methods on Massachusetts dataset. In the table, we can see that our method achieved a quality score of 0.65059, which is the highest in all the compared methods. The original U-Net, SegNet, PSPNet-50, Residual U-Net, DeepLabV3 and DANet obtain quality scores of 0.64873, 0.62477, 0.6271, 0.64271, 0.6141 and 0.6334, respectively. We think the superior performance of our method is due to the more powerful reconstruction ability.

In the third part of this section, we compared our method with other five methods on Shaoshan dataset. The five methods include Zang et al. [27], Residual U-Net, PSPNet-50, ESPNet[70] and Chen et al. [1]. For these five methods, we just follow the results in [1].

Table IV shows the comparison results among our method and other five methods on Shaoshan dataset. From the table, we can see that our method also achieves the highest quality score in all the methods, obtaining a quality score as high as 0.7328. Compared with other five methods, we improve the performance about 14%, 4%, 7%, 6% and 2%, respectively. This experimental results further prove the good performance of our method.

Table III. THE COMPARISON RESULTS AMONG OUR METHOD AND OTHER SIX STATE-OF-THE-ART SEGMENTATION METHODS TESTED ON MASSACHUSETTS DATASET.

Method	Completeness	Correctness	Quality
U-Net[9]	0.76628	0.80876	0.64873
SegNet[66]	0.72053	**0.82459**	0.62477
PSPNet-50[10]	0.76261	0.77921	0.6271
Residual U-Net[67]	**0.79688**	0.76862	0.64271
DeepLabV3[68]	0.73984	0.78322	0.6141
DANet[69]	0.74218	0.81209	0.6334
Ours	0.78525	0.79141	**0.65059**

Table IV. THE COMPARISON RESULTS AMONG OUR METHOD AND OTHER FIVE METHODS TESTED ON SHAOSHAN DATASET.

Method	Completeness	Correctness	Quality
Zang et al. [27]	0.7786	0.7135	0.5963
ResidualUnet[29]	0.7454	0.9149	0.6970
PSPNet[10]	0.6888	0.9434	0.6615
ESPNet[70]	0.7431	0.8882	0.6795
Chen et al [1]	**0.8247**	0.8443	0.7159
Ours	0.7605	**0.9526**	**0.7328**

In the final part of this section, we analyze the effectiveness of reconstruction bias layers. Fig. 6 shows the comparison results of our method with different reconstruction bias layers tested on LRSNY dataset. U-Net with our first one upsampling layer means that the model only use the first enforced upsampling layer to replace the original first upsampling layer in original U-Net. From Fig. 6, we can see that the better performance is the more enforced upsampling layers are used. This experiment proves the effectiveness of our reconstruction bias strategy.

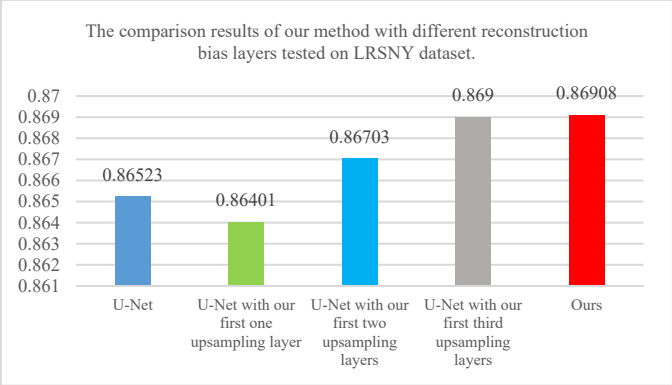


Fig. 6. The comparison results of our method with different reconstruction bias increased layers tested on LRSNY dataset.

CONCLUSION

In this paper, we proposed a reconstruction bias U-Net for road extraction from high resolution optical remote sensing images. Our reconstruction bias U-Net consisted of two parts: Encoding and Decoding. In the Encoding part, we used five operation couples of convolution, ReLU and max pooling. In the fourth and fifth operation couples, the drop out was also applied. In the Decoding part, we used four layers for upsampling. In each upsampling layer, multiple upsampling operation couples, which containing upsampling, convolution and ReLU, were set up. Every upsampling filter size was fixed at 2×2, but the convolutional filter sizes are different among one upsampling layer, obtaining multiple reconstruction information. Each upsampling layer was followed by an operation couple of concatenation, convolution, ReLU, convolution and ReLU. The final of our network was a sigmoid layer. The input and output sizes were 256×256×3 and 256×256×1, respectively. Besides, we proposed a publicly available dataset about road extraction from remote sensing images, called LRSNY. In the experimental part, we compared our method with other six state-of-the-art image segmentation method on LRSNY. Experimental results showed the good performance of our method and proved the effectiveness of our reconstruction bias part in our model. To further verify the performance of our model, we also compared our method with other methods on another two datasets: Massachusetts and Shaoshan Datasets. The experimental results on the two datasets both prove the effectiveness of our reconstruction bias model, as our model achieved the best performance among the compared methods on the both two datasets.

ACKNOWLEDGEMENTS

This study was financially supported by National Natural Science Foundation of China (No. 62001175), Natural Science Foundation of Fujian Province (No.2019J01081), United National Natural Science Foundation of China (No.U1605254), National Natural Science Foundation of China (No. 6187606, 61972167 and 61673186), and the Special National Key Research and Development Plan (No. 2019YFC1604705).

REFERENCES

[1] Z. Chen, W. Fan, B. Zhong, J. Li, J. Du, and C. Wang, "Corse-to-fine road extraction based on local Dirichlet mixture models and multiscale-high-order deep learning," *IEEE Transactions on Intelligent Transportation Systems,* vol. 21, no. 10, pp. 4283 - 4293, 2020.

[2] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 6, pp. 3322-3337, 2017.

[3] M. Maboudi, J. Amini, S. Malihi, and M. Hahn, "Integrating fuzzy object based image analysis and ant colony optimization for road extraction from remotely sensed images," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 138, pp. 151-163, 2018.

[4] M. O. Sghaier, and R. Lepage, "Road Extraction From Very High Resolution Remote Sensing Optical Images Based on Texture Analysis and Beamlet Transform," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* vol. 9, no. 5, pp. 1946-1958, 2017.

[5] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 130, pp. 139-149, 2017.

[6] I. Coulibaly, N. Spiric, R. Lepage, and M. St-Jacques, "Semiautomatic road extraction from VHR images based on multiscale and spectral angle in case of earthquake," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 11, no. 1, pp. 238-248, 2017.

[7] Y. Zang, C. Wang, Y. Yu, L. Luo, K. Yang, and J. Li, "Joint enhancing filtering for road network extraction," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 3, pp. 1511-1525, 2016.

[8] Q. Guo, and Z. Wang, "A Self-Supervised Learning Framework for Road Centerline Extraction From High-Resolution Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 13, pp. 4451-4461, 2020.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." pp. 234-241.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6230-6239.

[11] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS--A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion," *arXiv preprint arXiv:1906.07789*, 2019.

[12] S. Mohajerani, and P. Saeedi, "Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery." pp. 1029-1032.

[13] X.-Y. Tong, G.-S. Xia, Q. Lu, H. Shen, S. Li, S. You, and L. Zhang, "Learning transferable deep models for land-use classification with high-resolution remote sensing images," *arXiv preprint arXiv:1807.05713*, 2018.

[14] I. Nigam, C. Huang, and D. Ramanan, "Ensemble knowledge transfer for semantic segmentation." pp. 1499-1508.

[15] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raska, "Deepglobe 2018: A challenge to parse the earth through satellite images." pp. 172-17209.

[16] M. Zhang, X. Hu, L. Zhao, Y. Lv, M. Luo, and S. Pang, "Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images," *Remote Sensing,* vol. 9, no. 5, pp. 500, 2017.

[17] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 11, pp. 6054-6068, 2017.

[18] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding." pp. 3213-3223.

[19] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters,* vol. 30, no. 2, pp. 88-97, 2009.

[20] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." pp. 3234-3243.

[21] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687,* vol. 2, no. 5, pp. 6, 2018.

[22] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes." pp. 4990-4999.

[23] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving." pp. 954-960.

[24] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark." pp. 3226-3229.

[25] V. Mnih, *Machine learning for aerial image labeling*: Citeseer, 2013.

[26] F. Bastani, S. He, S. Abbar, M. Alizadeh, and H. Balakrishnan, "RoadTracer: Automatic Extraction of Road Networks from Aerial Images."

[27] Y. Zang, C. Wang, Y. Yu, L. Luo, K. Yang, and J. Li, "Joint Enhancing Filtering for Road Network Extraction," *IEEE Transactions on Geoscience & Remote Sensing,* vol. 55, no. 3, pp. 1511-1525, 2017.

[28] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network," *IEEE Transactions on Geoscience & Remote Sensing,* vol. 55, no. 6, pp. 3322-3337, 2017.

[29] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geoscience & Remote Sensing Letters,* vol. PP, no. 99, pp. 1-5, 2017.

[30] M. Maboudi, J. Amini, S. Malihi, and M. Hahn, "Integrating fuzzy object based image analysis and ant colony optimization for road extraction from remotely sensed images," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 138, pp. 151-163, 2018.

[31] R. Alshehhi, P. R. Marpu, L. W. Wei, and M. D. Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 130, pp. 139-149, 2017.

[32] I. Coulibaly, N. Spiric, R. Lepage, and M. St-Jacques, "Semiautomatic Road Extraction From VHR Images Based on Multiscale and Spectral Angle in Case of Earthquake," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* vol. PP, no. 99, pp. 1-11, 2017.

[33] Z. Lv, Y. Jia, Q. Zhang, and Y. Chen, "An Adaptive Multifeature Sparsity-Based Model for Semiautomatic Road Extraction From High-Resolution Satellite Images in Urban Areas," *IEEE Geoscience & Remote Sensing Letters,* vol. PP, no. 99, pp. 1-5, 2017.

[34] M. Li, A. Stein, W. Bijker, and Q. Zhan, "Region-based urban road extraction from VHR satellite images using Binary Partition Tree," *International Journal of Applied Earth Observation & Geoinformation,* vol. 44, pp. 217-225, 2016.

[35] D. Yin, S. Du, S. Wang, and Z. Guo, "A Direction-Guided Ant Colony Optimization Method for Extraction of Urban Road Information From Very-High-Resolution Images," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* vol. 8, no. 10, pp. 4785-4794, 2016.

[36] C. Poullis, "Tensor-Cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 95, no. 95, pp. 93-108, 2014.

[37] Y. Ren, Y. Yu, and H. Guan, "DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery," *Remote Sensing,* vol. 12, no. 18, pp. 2866, 2020.

[38] C. Tao, J. Qi, Y. Li, H. Wang, and H. Li, "Spatial information inference net: Road extraction using road-specific contextual information," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 158, pp. 155-166, 2019.

[39] G. Cheng, F. Zhu, S. Xiang, and C. Pan, "Road Centerline Extraction via Semisupervised Segmentation and Multidirection Nonmaximum Suppression," *IEEE Geoscience & Remote Sensing Letters,* vol. 13, no. 4, pp. 545-549, 2017.

[40] G. Cheng, F. Zhu, S. Xiang, Y. Wang, and C. Pan, "Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting," *Neurocomputing,* vol. 205, no. C, pp. 407-420, 2016.

[41] Y. Zang, C. Wang, L. Cao, Y. Yu, and J. Li, "Road Network Extraction via Aperiodic Directional Structure Measurement," *IEEE Transactions on Geoscience & Remote Sensing,* vol. 54, no. 6, pp. 3322-3335, 2016.

[42] Z. Hui, Y. Hu, S. Jin, and Z. Y. Yao, "Road centerline extraction from airborne LiDAR point cloud based on hierarchical fusion and optimization," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 118, pp. 22-36, 2016.

[43] L. Courtrai, and S. Lefèvre, "Morphological Path Filtering at the Region Scale for Efficient and Robust Road Network Extraction from Satellite Imagery," *Pattern Recognition Letters,* vol. 83, pp. 195-204, 2016.

[44] R. Liu, J. Song, Q. Miao, P. Xu, and Q. Xue, "Road centerlines extraction from high resolution images based on an improved directional segmentation and road probability," *Neurocomputing,* vol. 212, no. C, pp. 88-95, 2016.

[45] W. Shi, Z. Miao, and J. Debayle, "An Integrated Method for Urban Main-Road Centerline Extraction From Optical Remotely Sensed Imagery," *IEEE Transactions on Geoscience & Remote Sensing,* vol. 52, no. 6, pp. 3359-3372, 2014.

[46] R. Liu, Q. Miao, J. Song, Y. Quan, Y. Li, P. Xu, and J. Dai, "Multiscale road centerlines extraction from high-resolution aerial imagery," *Neurocomputing,* vol. 329, pp. 384-396, 2019.

[47] W. Shi, Z. Miao, Q. Wang, and H. Zhang, "Spectral–Spatial Classification and Shape Features for Urban Road Centerline Extraction," *IEEE Geoscience & Remote Sensing Letters,* vol. 11, no. 4, pp. 788-792, 2014.

[48] S. Movaghati, A. Moghaddamjoo, and A. Tavakoli, "Road Extraction From Satellite Images Using Particle Filtering and Extended Kalman Filtering," *IEEE Transactions on Geoscience & Remote Sensing,* vol. 48, no. 7, pp. 2807-2817, 2010.

[49] S. Leninisha, and K. Vani, "Water flow based geometric active deformable model for road network," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 102, pp. 140-147, 2015.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778.

[51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in International Conference on Neural Information Processing Systems, 2012, pp. 1097-1105.

[52] J. Redmon, and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017, pp. 6517-6525.

[53] R. Girshick, "Fast R-CNN," in The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448.

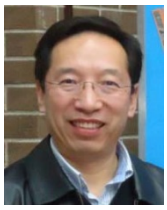
[54] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in International Conference on Neural Information Processing Systems, 2015, pp. 91-99.

[55] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708.

[56] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in ICLR, 2015.
[57] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1-9.
[58] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. PP, no. 99, pp. 1-1, 2017.
[59] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2261-2269.
[60] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 4, pp. 2043-2056, 2018.
[61] X. Lu, Y. Zhong, Z. Zheng, Y. Liu, J. Zhao, A. Ma, and J. Yang, "Multi-scale and multi-task deep learning framework for automatic road extraction," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 11, pp. 9362-9377, 2019.
[62] L. Gao, W. Song, J. Dai, and Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network," *Remote Sensing,* vol. 11, no. 5, pp. 552, 2019.
[63] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction from High-Resolution Remote Sensing Data," *IEEE Access,* vol. 8, pp. 179424 - 179436, 2020.
[64] X. Yang, X. Li, Y. Ye, R. Y. Lau, X. Zhang, and X. Huang, "Road detection and centerline extraction via deep recurrent convolutional neural network u-net," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 9, pp. 7209-7220, 2019.
[65] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234-241.
[66] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 12, pp. 2481-2495, 2017.
[67] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters,* vol. 15, no. 5, pp. 749-753, 2018.
[68] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 4, pp. 834-848, 2018.
[69] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019, pp. 3146-3154.
[70] M. Sachin, R. Mohammad, C. Anat, S. Linda, and H. Hannaneh, "ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation," in IEEE Conference on European Conference on Computer Vision(ECCV), 2018.

Ziyi Chen received his Ph.D. degree in signal and information processing from Xiamen University, China, in 2016. He is currently a lecturer in the Department of Computer Science and Technology, Huaqiao University, China. His current research interests include computer vision, machine learning, and remote sensing image processing.

Cheng Wang (M'12) received the Ph.D. degree in information communication engineering from the National University of Defense Technology, China, in 2002. He is currently a Professor at the School of Information Science and Engineering, Xiamen University, China. His current research interests include remote sensing image processing, mobile laser scanning data analysis, and multi-sensor fusion.



Jonathan Li (M'00-SM'11) received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa, in 2000. He is currently a Professor at the Department of Geography and Environmental Management, University of Waterloo, Canada. His current research interests include information extraction from earth observation images and 3-D surface reconstruction from mobile laser scanning point clouds.

Nianci Xie is currently a junior student of Huaqiao University. His major is computer science. His research interests is computer vision.

Yan Han is currently a junior student of Huaqiao University. His major is computer science. His research interests is computer vision.

Jixiang Du received the B.Sc. and M.Sc. degrees in vehicle engineering from the Hefei University of Technology, Hefei, China, in 1999 and 2002, respectively, and the Ph.D. degree in pattern recognition and intelligent system with the University of Science and Technology of China, Hefei, in 2005. He is currently a Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China.