Corse-to-Fine Road Extraction Based on Local Dirichlet Mixture Models and Multiscale-High-Order Deep Learning

Ziyi Chen, *Member, IEEE*, Wentao Fan, *Member, IEEE*, Bineng Zhong, *Member, IEEE*, Jonathan Li, *Senior Member, IEEE*, Jixiang Du, *Member, IEEE*,
and Cheng Wang, *Senior Member, IEEE*

***Abstract*— Road extraction from remote sensing images is an attractive but difficult task. Gray-value distribution and structure feature information are both crucial for road extraction task. However, existing methods mainly focus on structure feature information which contains morphological shape features and machine learning features, suffering from lots of false positives which are generated at positions having similar structure features but different gray-value distribution with roads. To effectively fuse the two complementary gray-value distribution and structure feature information, we propose a coarse-to-fine road extraction algorithm from remote sensing images. First, at the coarse level, we introduce a local Dirichlet mixture models (LDMM) which utilizing gray-value distribution information to pre-segment images into potential roads and backgrounds. Thus, most backgrounds having different gray-value distribution with roads can be removed firstly. Compared with original Dirichlet mixture models, the LDMM is much faster and more accurate. Next, at the fine level, we introduce a multiscal-high-order deep learning strategy based on ResNet model which can learn robust structure context features for final road extraction step. Based on the results of LDMM, the multiscal-high-order strategy can further remove false positives which have different structure features with roads. Compared with a single scanning size ResNet, our multiscale-high-order strategy can learn higher-order context information, leading to better performances. We test our algorithm on Shaoshan dataset. Experiments illustrate our better performance compared with other six state-of-the-art methods.**

***Index Terms*— Road extraction, remote sensing image, local Dirichlet mixture model, multiscal-high-order, deep learning.**

I. Introduction

AUTOMATIC road extraction from remote sensing images has become more and more important since it is an essential preprocessing step for various applications, such as navigation, road network planning, road network information update, etc. Compared with traditional manual road areas labeling, automatic road extraction from remote sensing images is less time consuming, more economic and effective [1]. Much research has focused on automatic road extraction from remote sensing images [1]–[10]. However, due to the challenges of noise, shadows, complexity of the background, occlusions (generated by trees, clouds, and buildings etc.) in raw remote sensing images, automatic road extraction from remote sensing images is still a difficult and attractive research topic.

Gray-value distribution and structure feature information are both crucial for road extraction from remote sensing images as remote sensing images usually cover large areas with complex backgrounds. Considering only gray-value distribution or structure feature information for road extraction may result to lots of false positives. The gray-value based methods may generate false positives where have similar gray-value distribution but different structure features with roads. On the other hand, the structure-based methods may generate false positives where have similar structure features but different gray-value distribution with roads. However, based on our studies, most existing methods for road extraction from remote sensing images mainly concerned structure feature information which contains morphological shape features and machine learning features. Thus, to effectively fuse two complementary gray-value distribution and structure feature information, we propose a coarse-to-fine road extraction algorithm from remote sensing images.

First, at the coarse level, a finite Dirichlet mixture is applied to pre-segment images into roads and backgrounds. A finite Dirichlet mixture model adopts the Dirichlet distribution as the parent distribution to model image pixel value probability density for image segmentation. Dirichlet

Manuscript received January 21, 2019; revised June 10, 2019; accepted August 28, 2019. This work was supported in part by the Natural Science Foundation of Fujian Province under Grant 2019J01081, in part by the Huaqiao University Foundation under Grant 600005-Z16X0115, in part by the Fujian Key Laboratory of Sensing and Computing for Smart Cities of Xiamen University, and in part by the National Natural Science Foundation of China under Grant 61876068, Grant 61572205, Grant U1605254, Grant 61972167, and Grant 61673186. The Associate Editor for this article was D. F. Wolf. *(Corresponding author: Cheng Wang.)*

Z. Chen, W. Fan, B. Zhong, and J. Du are with the Department of Computer Science and Technology, Huaqiao University, Xiamen 361021, China (e-mail: chenziyihq@hqu.edu.cn; fwt@hqu.edu.cn; bnzhong@hqu.edu.cn; jxdu@hqu.edu.cn).

J. Li is with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology (MOE), Xiamen University, Xiamen 361005, China, and also with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@xmu.edu.cn).

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

Digital Object Identifier 10.1109/TITS.2019.2939536

mixture model achieved good performances in normal image segmentation [11] and data clustering [12]. However, when directly applying Dirichlet mixture model for segmentation of remote sensing images, it obtains poor performance due to pixel gray-values' inhomogeneity within roads (or backgrounds). Besides, original Dirichlet mixture model has high computational complexity when processing an image with a large size. To overcome the above problem, we propose a local Dirichlet mixture model (LDMM) to pre-segment images into roads and backgrounds. Compared with original Dirichlet mixture model, LDMM is much faster and more accurate. Through LDMM, most backgrounds having different gray-value distribution with roads are removed.

Second, at the fine level, we propose a multiscale-high-order deep learning strategy (MTL) to further remove false positives on the results of LDMM. CNN model has been proved to be an excellent model for learning structure features, as it performs superior to other types of methods in many computer vision research areas. However, one single scanning size of CNN model is unable to catch high-order context information which is important to deal with remote sensing image processing under complex backgrounds. Thus, we propose multiscale-high-order deep learning strategy to catch high-order context information for road extraction from remote sensing images. In MTL, we use patches with different sizes to train different CNN classifiers (which are ResNet-50 classifiers in our method). After that, we merge results extracted from ResNet-50 classifiers with different scanning sizes to retrain a high-order ResNet-50 classifier, based on which we can extract the high-order context feature and obtain the final road extraction results.

The major contributions of this paper lie on:

(1) This paper proposes a coarse-to-fine strategy, which considers both pixel gray-values' distribution and mulitiscale-high-order features, for road extraction from high-resolution remote sensing images.

(2) This paper proposes a local Dirichlet mixture model. compared with original Dirichlet mixture model, the proposed LDMM is much faster while achieving a higher precision.

(3) This paper proposes a higher-order feature extraction method based on the multi-scale feature extraction results of ResNet. The experimental results show that the proposed higher-order feature extraction method is more robust and can obtain a better performance compared with state-of-the-art methods.

II. Related Work

Generally speaking, the road extraction task contains two subtasks: road area extraction and road centerline extraction [1], [7]. Road area extraction methods produce pixel-level labeling of roads [2]–[5], [7], [9], [10], [13], [14], while skeletons of roads are extracted for road centerline extraction [1], [6], [8], [15]–[20].

For road centerline extraction, morphological thinning algorithm [21], regression-based method [20], and nonmaximum suppression-based method [16] have been widely used.

Road area extraction, which is the focuses of this paper, can be considered as a segmentation or pixels-level classification problem [7]. During road area extraction, the morphological and machine learning features are widely used [1], [3], [4], [22]. Feature extraction and sample selection strategy are also widely used in other hot research areas [23]–[26].

In the following two sections, we present a detailed review of existing methods for road extraction from remote sensing images and remote sensing applications using CNN.

A. Studies on Road Extraction From Remote Sensing Images

Song et al. combined shape index feature and support vector machine (SVM) to extract road areas. Movaghati et al. proposed a road extraction method from satellite images using particle filtering (PF) and extended kalman filtering (EKF) [27]. The PF is combined with EKF to find best continuation of the road after an obstacle or junction, which achieved satisfactory results. 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 [10]. 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 [28]. Grinias et al. proposed a novel segmentation algorithm based on Markov random field model. The key point of their method lies on the class-driven vector data quantization and clustering. Finally, the Random Forest was applied to obtain a good classification rate [29]. Cheng proposed a road region extraction method by incorporating multiple features and multiscale fusion, which achieved satisfactory visual performances compared with other methods [16]. Liu et al. proposed a road network extraction framework. In the first stage, they combined shear transform with directional segmentation to get the initial road regions [19]. Troya-Galvis et al. proposed an approach which combined two different approaches for automatic remote sensing image pixel-level interpretation. They obtained satisfactory results when applied for road extraction from remote sensing images [30]. Zang et al. proposed a novel aperiodic directional structure measurement (ADSM) for road network extraction. Through ADSM, they well characterized road-like structures which can be used as the guidance to construct a mask to denote potential road regions. Experimental results demonstrated their method's effectiveness and efficiency [8]. Zang et al. also proposed a joint enhancing filtering framework to generate a pre-processed image for the road network extraction in the next stage [6]. 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 [13].

B. Remote Sensing Applications Using CNN

In recent years, deep convolutional neural networks (CNN) have led a series of breakthroughs for computer vision tasks [31]–[43]. Maggiori et al. used CNN for large-scale remote-sensing image classification [44]. To address the issue of imperfect of training data, they separated the training into two stages: training with large amounts of possibly inaccurate reference data; refining with a small amount of accurately labeled data. They achieved rather good results in the experiments. Zhao et al. proposed an algorithm, called multiscale CNN, to learn spatial-related deep features for classifying remotely sensed imageries [45]. They used multiscale CNN to learn high-level spatial features by using the hierarchical learning structure and capture contextual information by using multiscal learning scheme. Alshehhi et al. proposed a single patch-based CNN for extraction of roads and buildings from high-resolution remote sensing data [4]. 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 [1]. Zhang et al. used a semantic segmentation neural network which combines the strengths of residual learning and U-Net [46] for road area extraction from remote sensing images [7]. They achieved better results compared with other state-of-the-arts approaches. Chen et al. proposed two frameworks which contained deep fully convolutional networks with shortcut blocks for semantic segmentation from very-high-resolution remote sensing images [47]. Zhuang proposed a Dense Relation Network for semantic segmentation and achieved good performance [48].

III. Method

In this section, we first give an introduction about the framework of our method. Then, a Dirichlet mixture model is introduced in order to improve the performance of the Dirichlet mixture model [11]. Third, we introduce the ResNet that we used for deep learning. Fourth, we propose multiscale-high-order deep learning strategy.

A. The Framework of Our Method

Our framework consists of two stages: training stage and extraction stage. Fig.1 and Fig.2 show the training and extraction frameworks of our method, respectively. Although Dirichlet mixture model does not need to pre-train, the ResNet used for final road area extraction needs to train the model parameters, fitting for our goal.

In the training stage, we first generate positive patches (road patches) and negative patches (background patches) with sizes s1 × s1, s2 × s2, s3 × s3, respectively. Then, we use the generated training patches to train three ResNet-50 models (each for a scanning size). To make ResNet fitting for our 2-class segmentation, we reconstruct the final fully connection layers. Next, we use the proposed LDMM to generate coarse 2-class segmentations. After that, the trained ResNet models are used to extraction road areas based on LDMM segmentation results. To obtain high-order information, extraction results of three ResNet-50 models are merged to generate training patches for final high-order ResNet-50 model.

In the extraction stage, we use LDMM to obtain a coarse road and background segmentation result. On the other side, we use the trained ResNet models with different scanning sizes to get road extraction results based on LDMM results. Then, we merge the extraction results of three different scanning size ResNet models. Finally, with merge results as inputs, we use the trained multiscale-high-order ResNet model to extract road areas based on LDMM results.

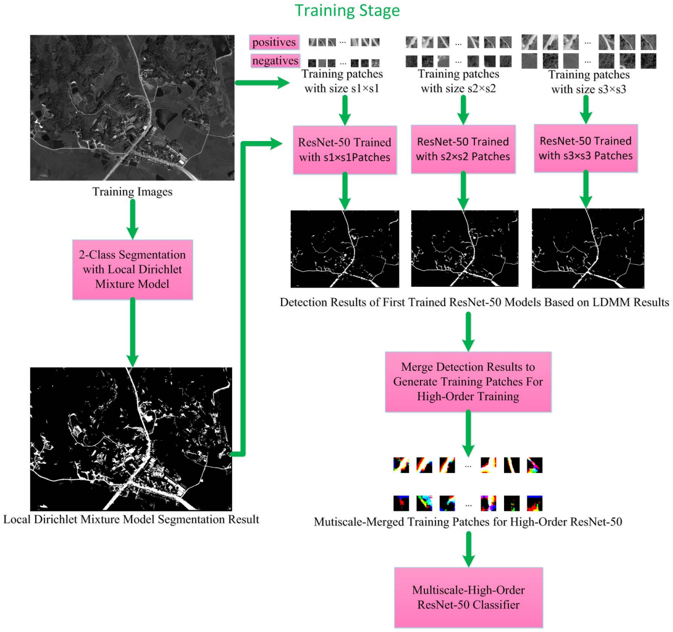

Training Stage
positives
negatives
Training patches with size s1×s1
Training patches with size s2×s2
Training patches with size s3×s3
Training Images
ResNet-50 Trained with s1×s1Patches
ResNet-50 Trained with s2×s2 Patches
ResNet-50 Trained with s3×s3 Patches
2-Class Segmentation with Local Dirichlet Mixture Model
Detection Results of First Trained ResNet-50 Models Based on LDMM Results
Merge Detection Results to Generate Training Patches For High-Order Training
Local Dirichlet Mixture Model Segmentation Result
Mutiscale-Merged Training Patches for High-Order ResNet-50
Multiscale-High-Order ResNet-50 Classifier

Fig. 1. The training stage of the proposed method.

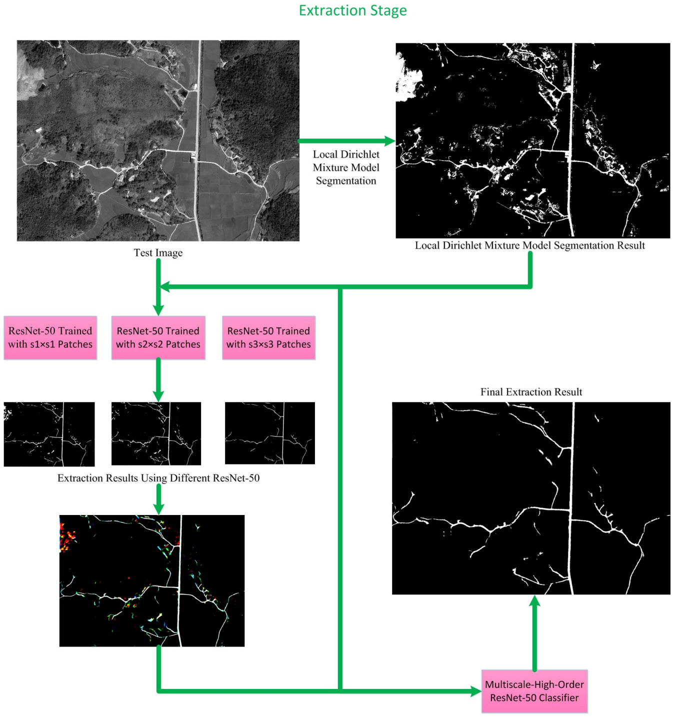

Extraction Stage
Local Dirichlet Mixture Model Segmentation
Test Image
Local Dirichlet Mixture Model Segmentation Result
ResNet-50 Trained with s1×s1 Patches
ResNet-50 Trained with s2×s2 Patches
ResNet-50 Trained with s3×s3 Patches
Final Extraction Result
Extraction Results Using Different ResNet-50
Multiscale-High-Order ResNet-50 Classifier

Fig. 2. Extraction Stage of the proposed method.

B. Local Dirichlet Mixture Models

In our framework, we follow the finite mixture of Dirichlet mixture model proposed in [11] as our basic Dirichlet mixture model.

For each pixel $\vec{\mathrm{X}}_i$, which is assumed to distribute according to a spatially constrained Dirichlet mixture models with M components, then its probability density function can be represented by:

$$\mathrm{p}(\vec{\mathrm{X}}_i|\vec{\pi},\vec{\mathrm{a}}) = \sum_{j=1}^{M} \pi_{ij} Dir(\vec{\mathrm{X}}_i|\vec{\mathrm{a}}_j), \tag{1}$$

where $\vec{\pi}_i = (\pi_{i1}, \ldots, \pi_{iM})$ denotes the mixing coefficients, which are positive and sum to one, $\sum_{j=1}^{M} \pi_{ij} = 1$, $\vec{a} = (\vec{a}_1, \ldots, \vec{a}_M)$, and $Dir(\vec{X}_i|\vec{a}_j)$ is the Dirichlet distribution of component j with its own positive parameters $\vec{a}_j = (a_{j1}, \ldots a_{jD})$:

$$Dir(\vec{X}_i|\vec{a}_j) = \frac{\Gamma(\sum_{l=1}^{D} a_{jl})}{\prod_{l=1}^{D} \Gamma(a_{jl})} \prod_{l=1}^{D} X_{il}^{a_{jl}-1}, \tag{2}$$

where $\vec{X}_i = (X_{i1}, \ldots, X_{iD})$, D is the dimensionality of $\vec{X}_i$ and $\sum_{l=1}^{D} X_{il} = 1$, $0 \le X_{il} \le 1$ for $l = 1, \ldots, D$.

Consider a set of N independent identically distributed vectors $\chi = \{\vec{X}_1, \ldots, \vec{X}_N\}$ assumed to be generated from the mixture distribution in Eq. (1). The likelihood function of the Dirichlet mixture model is given by:

$$p(\chi|\vec{\pi},\vec{a}) = \prod_{i=1}^{N} \{\sum_{j=1}^{M} \pi_{ij} Dir(\vec{X}_i|\vec{a}_j)\}. \tag{3}$$

For each vector $\vec{X}_i$ introducing a M-dimensional binary random vector $\vec{Z}_i = \{Z_{i1}, \ldots, Z_{iM}\}$, such that $Z_{ij} \in \{0, 1\}$, $\sum_{j=1}^{M} Z_{ij} = 1$, and $Z_{ij} = 1$ if $\vec{X}_i$ belong to component j and 0, otherwise. For the latent variables $Z = \{\vec{Z}_1, \ldots, \vec{Z}_N\}$, the conditional distribution of Z given the mixing coefficients $\vec{\pi}$ is defined as:

$$p(Z|\vec{\pi}) = \frac{K!}{\prod_{j=1}^{M} (Z_{ij})!} \prod_{j=1}^{M} \pi_{ij}^{Z_{ij}}. \tag{4}$$

To impose local spatial smoothness between adjacent pixels into mixture model, each pixel in an image with the average value of its neighbors (includes itself) can be described as:

$$\vec{Z}_{ij} = \frac{\sum\limits_{m \in \Omega_i} Z_{mj}^{(t-1)}}{|\Omega_i|}, \tag{5}$$

where Ω_i denotes the neighborhood of the ith pixel, $|\Omega_i|$ is the number of pixels in the neighborhood of the ith pixel, (t-1) indicates the iteration of the previous step. The prior distribution of $\vec{\pi}_j$ follows a Dirichlet distribution as follow:

$$p(\vec{\pi}_j) = Dir(\vec{\pi}_j|\vec{\Lambda}_i) = \frac{\Gamma(\sum_{j=1}^{M} a_k^2 \vec{Z}_{ij}^b)}{\Pi_{j=1}^{M} \Gamma(a_k^2 \vec{Z}_{ij}^b)} \prod_{j=1}^{M} \pi_{ij}^{a_k^2 \vec{Z}_{ij}^b - 1}, \tag{6}$$

where the Dirichlet parameter $\vec{\Lambda}_i = a_k^2 \vec{Z}_{ij}^b$.

Finally, to estimate the model parameters, we follow the variational inference learning algorithm as developed in [11]. Readers can get the details from reference [11].

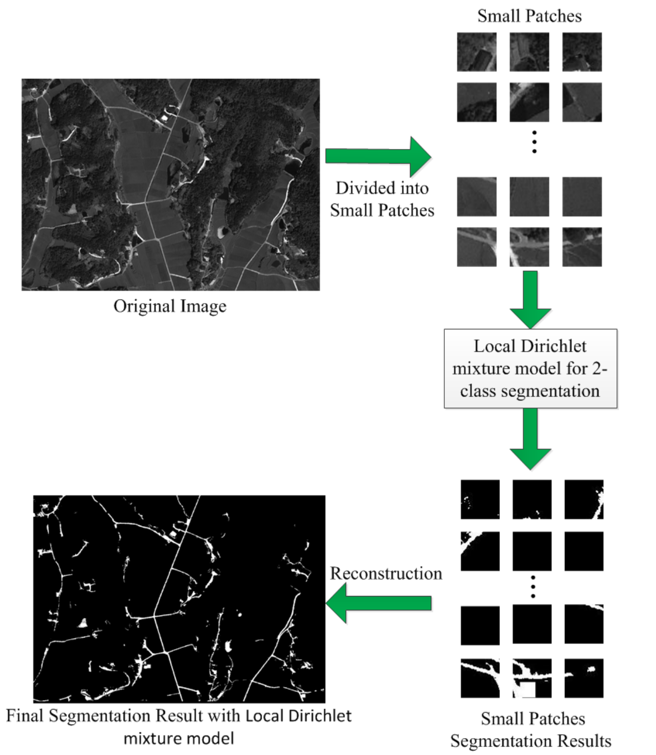

Small Patches
Divided into Small Patches
Original Image
Local Dirichlet mixture model for 2-class segmentation
Reconstruction
Final Segmentation Result with Local Dirichlet mixture model
Small Patches Segmentation Results

Fig. 3. The framework of local Dirichlet mixture model.

After parameter estimation, the finite mixture of Dirichlet mixture model can be directly applied to road and background segmentation on remote sensing images. However, remote sensing images usually cover large areas of ground, and the Dirichlet mixture model often fails to separate road from background (or separating background from road), as shown by the following Fig.7. The reason for this phenomenon is that the original Dirichlet mixture model needs to estimate model parameters by road and background pixel gray-value distributions while road and background pixel gray-values are strongly overlapped within remote sensing images. Thus, the essential problem of directly applying the original Dirichlet mixture model for road extraction from remote sensing image is using parameters estimated globally.

Note that although road and background pixel gray-values have many overlaps in a large-size remote sensing image, it will have much fewer overlaps in a local small patch of remote sensing image. Thus, we use a local Dirichlet mixture model instead of the original Dirichlet mixture model. The framework of local Dirichlet mixture model is shown in Fig. 3. The detail steps of our local Dirichlet mixture model are described as follows:

(1) Assume a given remote sensing image I_m with a size of $M \times N$, we divide the I_m into K small patches p_a with a size of $a \times a$, a is much smaller than M or N.

(2) For each patch p_{ai}, we segment the patch into 2 classes by Dirichlet mixture model and save the segmentation result.

(3) Re-built the segmentation result of whole image by merging segmentation results from all patches.

C. ResNet for Deep Learning

The ResNet is proposed by He *et al.* [31]. In ResNet, to overcome the degradation problem of deep networks, the

TABLE I
THE NETWORK OF RESNET-50 USED IN OUR METHOD

layer name	output size	ResNet-50-layer
conv1	112×112	7×7 ,64, stride 2 3×3 max pool, stride 2
conv2_x	56×56	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
onv3_x	28×28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$
conv4_x	14×14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$
conv5_x	7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
fc_x	1×1	average pool, 1000-d fc, 500-d fc, 2-d fc

residual learning framwork is introduced. Considering H(x) as an underlying mapping to be fit by a few stacked layers, with x denoting the inputs to the first of these layers. The original approximation mapping function can be $H(x) = F(x) + x$ when let $F(x) := H(x) - x$ (Here $F(x)$ is the residual item). In ResNet, residual learning is applied to every few stacked layers.

Assuming x and y are the input and output vectors of the layers considered, a building block can then be defined as:

$$y = F(x, \{W_i\}) + x. \tag{7}$$

The function $F(x, \{W_i\})$ represents the residual mapping to be learned. When a block has two layers, the residual mapping can be $F = W_2\sigma(W_1 x)$ where σ denotes the output of the second layer. The operation $F + x$ is performed by a shortcut connection and element-wise addition.

In our method, we use a ResNet-50 as the CNN network of our deep learning strategy. Table I shows the network framework of ResNet-50 used in our method. It can be seen from table I that ResNet-50 contains 49 convolution layers and a max pooling layer before the final fully-connection layers. For these layers, we follow the original settings of He's [31]. Note that, there are Relu and pooling layers following each convolution block, which are not showing in table I. Furthermore, we follow the original setting about the short cut connections. To make ResNet suitable for our situation, we replace the original fully connection layers with 3 fully connection layers, having 1000, 500 and 2 neurons, respectively.

D. Multiscale-High-Order Deep Learning Strategy

In the training stage, we train three different ResNet classifiers with different scanning sizes (i.e. the input patch sizes of ResNet).Then, using three classifiers to extract classification results pixel by pixel. The extraction results of different scanning size ResNet models are merged by channels to train a multiscale-higher-order ResNet model for final road extraction. The merge processing is shown in Fig.2.

In the test stage, the test image is first segmented into two classes by the proposed LDMM. Then, the test image is scanned by three ResNet models with different scanning sizes. Note that, the scanning process is based on the segmentation result of LDMM. Only the areas which are segmented into potential positives are scanned. After scanned by three different scanning size ResNet models, the extraction results are merged. Finally, using the merge results as inputs, the trained multiscale-high-order ResNet model is applied to extract road areas based on LDMM result. Fig.4 shows the flowchart of our proposed multiscale-high-order deep learning.

IV. RESULTS AND DISCUSSION

In this section, we give an introduction about dataset used in experiments at first. Then, detailed analyses about proposed LDMM are given. Finally, the experimental results and comparisons are presented.

A. Dataset

In our experiments, we verify and analyze the performances of the proposed method on ShaoShan dataset, which is a Pleiades optical remote sensing image of part ShaoShan (in China) with a resolution of 0.5m. The original image is a large image with size of 11125 × 7918. We divide the whole large image into 49 pieces, each of which has a size of 1589 × 1131. Among these images, we select 29 images as training images. And the remaining 20 images are used as test images. Fig. 5 shows several example images in our dataset. The first and second rows are training images and their corresponding labels, respectively. The third and fourth rows are test images and their corresponding labels, respectively. Note that the labels of images are generated by manual works.

B. Local Dirichlet Mixture Model Analyses

The goal of the experiments described in this section is to evaluate the effectiveness of the proposed local Dirichlet mixture model. To exhibit the superior of LDMM, we evaluate the performances of original Dirichlet mixture model [11] and proposed LDMM from two aspects: processing speed and correctness.

1) Processing Speed: In theoretical analysis, the computation complexities of original Dirichlet mixture model and LDMM are both approximate to O(n^2) × k for processing an image with size n × n, where k represents the iteration numbers. The difference between the original Dirichlet mixture model and LDMM is that the LDMM divides the processing image into many small patches with size m × m (m is much smaller than n). Thus, LDMM requires much less time than the original Dirichlet mixture model does. For example, when an image with size 1000 × 1000 is given, the computational complexity of original Dirichlet mixture model will be O(10^6) × 150, supposing the number of iterations is 150.

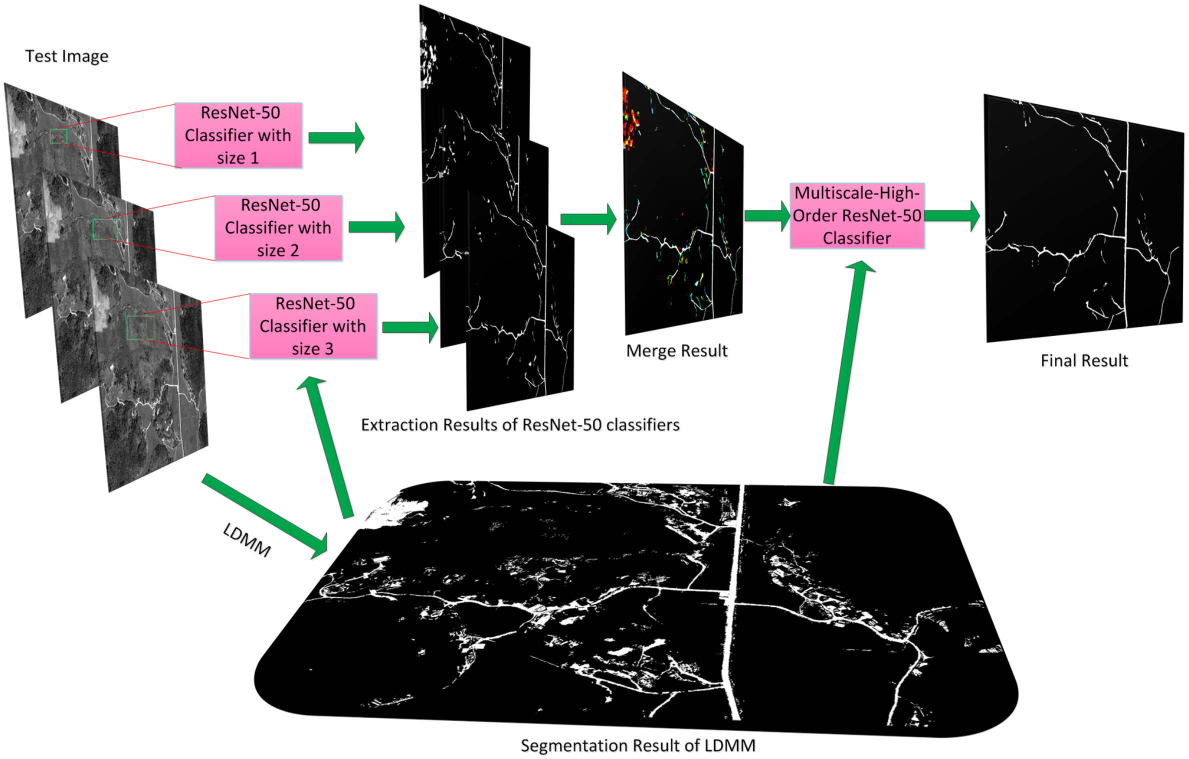

Test Image
ResNet-50 Classifier with size 1
ResNet-50 Classifier with size 2
ResNet-50 Classifier with size 3
Extraction Results of ResNet-50 classifiers
Merge Result
Multiscale-High-Order ResNet-50 Classifier
Final Result
LDMM
Segmentation Result of LDMM

Fig. 4. The illustration of multiscal-high-order deep learning.

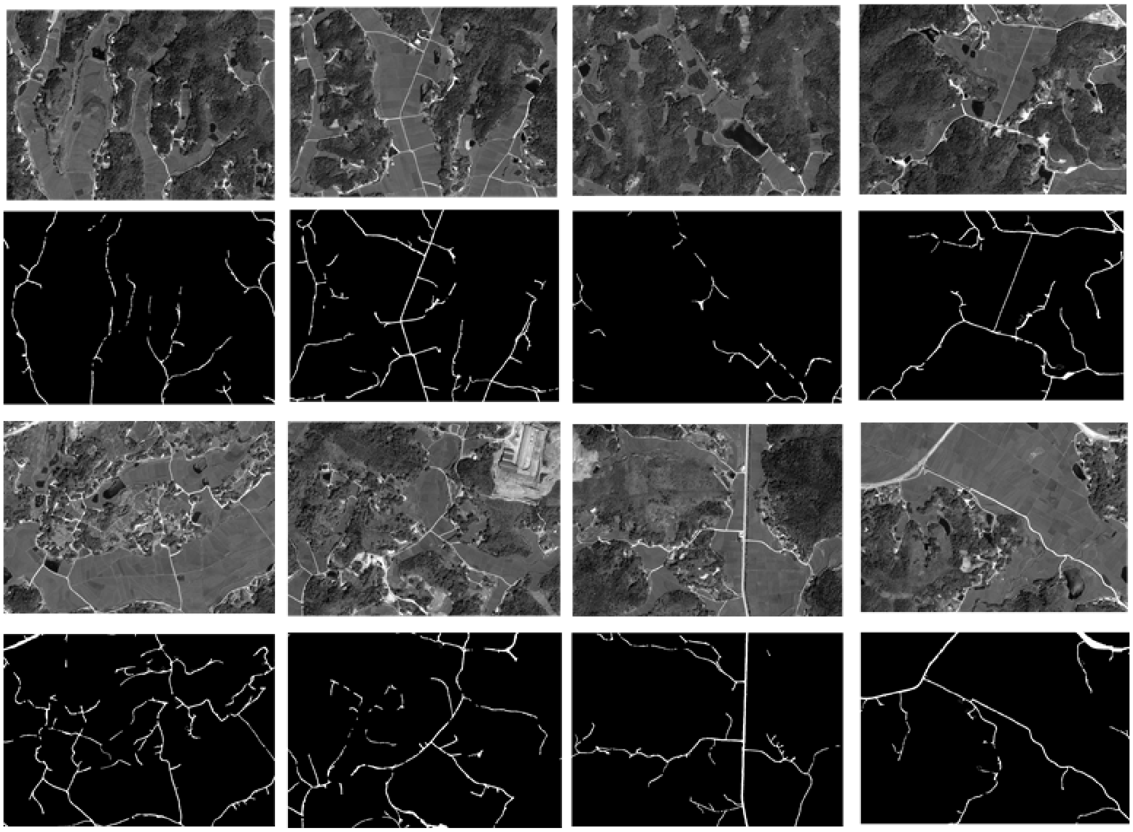


Fig. 5. Several examples of training, test, and corresponding label. Row 1 and 2 are training images and their corresponding labels, respectively. Row 3 and 4 are test images and their corresponding labels, respectively.

Suppose LDMM divides the test image into patches with size 20 × 20, then the computation complexity of LDMM will be 2500 (patch number) × O($10^2 \times 150$), which is lower than the original Dirichlet mixture model. Fig.6 shows the comparison of processing time per test image between original Dirichlet mixture model and the proposed LDMM. The experiments

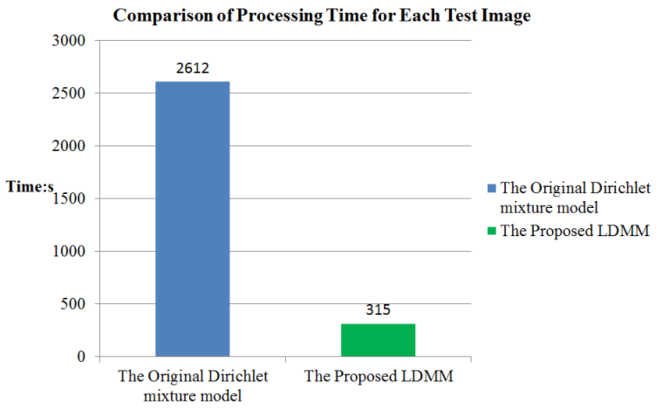


Fig. 6. Comparison of Processing Time for each test image between the original Dirichlet mixture model and the proposed LDMM.

run on a computer with Intel®Core™ i5-4570 CPU and 8GB memory. In our experiments, we set the size of the divided patches as 20 × 20. As seen from this figure, the average cost time for each test image of our LDMM is 315 seconds, while the original Dirichlet mixture model takes 2612 seconds (which is about 8 times of LDMM). This proves that the time complexity of our LDMM is much less than the original Dirichlet mixture model. A tip of LDMM is that we statistic the training images to learn the distribution of positives. With the learned distribution, each pixel is subtracted with positives' lowest gray value when applying LDMM on test images.

2) Correctness: To further illustrate the superior of LDMM to the original Dirichlet mixture model, we also compare the correctness of segmentation in test images in visualization and quantitation. Fig. 7 shows the visualization comparison of segmentation results of original Dirichlet mixture model and LDMM. Rows 1, 2 and 3 are original test images, segmentation results by original Dirichlet mixture model, and segmentation results by LDMM, respectively. Obviously, due to the complexity of pixel value distribution within one class in remote sensing images, the original Dirichlet mixture model performs poorly in test images. In contrast, as we use small region for Dirichlet mixture model segmentation in remote sensing image, our LDMM performs quite well. The following road extraction can benefit from LDMM in both processing time and accuracy.

We also give quantitative comparison between original Dirichlet mixture model and LDMM on test images. Table II gives the correctness comparison of original Dirichlet mixture model and the LDMM on Shaoshan test images. As shown in table II, the correctness of original Dirichlet mixture model is rather low while LDMM's is much better. Note that, most of the false segmentations are false positives. The correctness evaluation criterion is given in the following Eq. (8).

C. Comparison and Results

In this section, we illustrate the implementation details of our approach in experiments, and present the comparisons of our method with other state-of-the-art methods.

1) Implementation Details of Our Method: We select three different scanning sizes of ResNet-50 as 20 × 20, 30 × 30 and 40 × 40. For each scanning size of ResNet-50, we generate 200,000 positive patches and 200,000 negative patches to train ResNet-50 model. During training, we allow to randomly rotate training patches among 0-180 degrees. We use Keras's SGD optimizer with setting parameters "momentum equal" to 0.9, "decay" equal to le-6 and "lr" equal to 0.001. The training batch size is set to 64, steps per epoch is set to 3000. The total training epochs is 10. Note that, the label of a patch in our implementaion is just same with the label of its center pixel. For example, if the center pixel of a patch is labeled as road, then the patch is labeled as road. Otherwise, the patch is labeled as background.

After training three ResNet-50 with scanning sizes of 20 × 20, 30 × 30 and 40 × 40, we apply the three models to extract road areas in 29 training images. Next, we merge the extraction results obtained by three ResNet-50 models. With the merged results, training patches for final multiscale-high-order ResNet-50 classifier are generated. To overcome the merge problem caused by different patch sizes, we up-sample the patches of sizes 20 × 20 and 30 × 30 to size 40 × 40. Note that, we use the 'nearest' interpolation method for up-sample operation. Here, we generate 200,000 positive

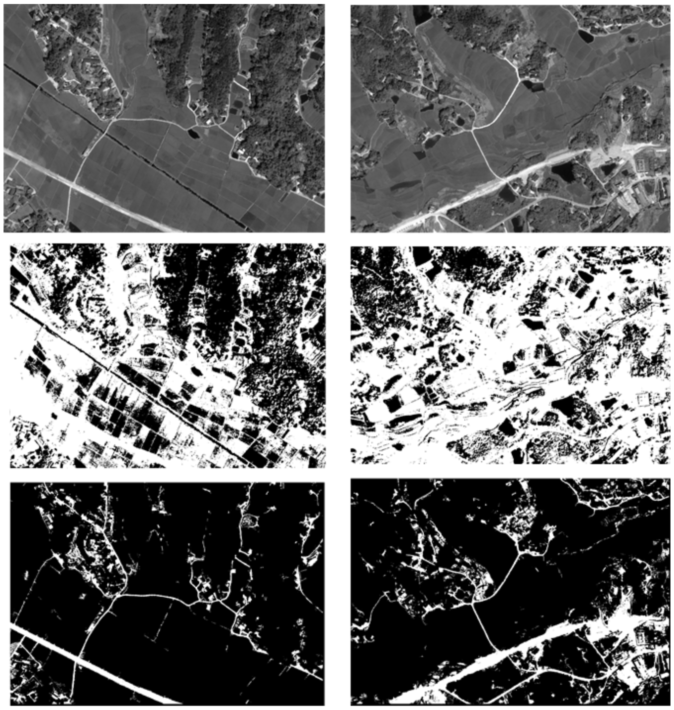


Fig. 7. The visual comparison of segmentation results by the original Dirichlet mixture model and the proposed LDMM. Row 1, 2 and 3 are original images, segmentation results of original Dirichlet mixture model and segmentation results of LDMM, respectively.

TABLE II
THE CORRECTNESS COMPARISON OF ORIGINAL DIRICHLET MIXTURE MODEL AND LDMM ON TEST IMAGES OF SHAOSHAN

Method	Correctness
Original Dirichlet mixture model[11]	0.0724
LDMM	0.3813

patches and 200,000 negative patches for final multiscale-high-order ResNet-50 training. During training patches generation, we also take label information and LDMM result into consideration. The detail training parameters are same as parameters used in training three different scanning sizes ResNet models.

During extraction stage, we first use LDMM to segment total 20 test images into 2 classes (potential roads and backgrounds). Based on the results of LDMM, we use the trained three ResNet-50 classifiers to scan the test images. Only the positions where are segmented into potential roads by LDMM are examined. Then, we can obtain three extraction results for each test image. For the final extraction step, we use the trained multiscale-high-order ResNet-50 to re-examine the test image based on LDMM result using merge results as inputs. For each scanning position, we merge the 20 × 20 (obtained by scanning size 20 × 20 ResNet-50), 30 × 30 (obtained by scanning size 30 × 30 ResNet-50) and 40 × 40 (obtained by scanning size 40 × 40 ResNet-50) area to generate the input for final examination of multiscale-high-order ResNet-50. After the above steps, we can obtain the final results. Fig. 8 shows several results of our method tested on Shaoshan dataset. The rows 1-4 are orginal images, labes, results of LDMM, final results of our method, respectively. As seen in Fig.8, our method obtains quite satisfactory results visually. Benefiting from the pre-segmentation by LDMM using gray-value distribution information, our results generate few false positives at positions where have similar structure feature but different gray-value distribution with roads, e.g. ridges in the farmland.

2) Quantitation Comparison: To quantitatively illustrate the performance of our method, we compare our method with three other state-of-the-art methods [6], [7], [31], [46], [49], [50] on Shaoshan dataset. For [6], as Zang et al. have tested their method on Shaoshan dataset, we reference their results directly. For [31], [46], we use the same training images as used in our method. In original ResNet, to suit for our 2 classes classification, we revise the final three fully-connection layers as same as RestNet-50 used in our method. In [46], we set Unet's input image size as 256 × 256. Thus, we divide the original 1589 × 1131 test images into 256 × 256 test images. After segmented by the trained Unet, we merge the 256 × 256 results to obtain the final 1589 × 1131 results. For [7], [49], [50], we use the original settings in the experiments except that the class number is set at 2 as we only have 2 classes. The codes are obtained from the GitHub.[1]

Table III shows the comparison results of [6], [31], [46], [7], [49], [50] and ours on Shaoshan dataset. The evaluation criteria are as follows:

$$\text{completeness} = \frac{TP}{TP+FN}$$
$$\text{correctness} = \frac{TP}{TP+FP},$$
$$\text{quality} = \frac{TP}{TP+FN+FP} \tag{8}$$

TABLE III
THE COMPARISON AMONG [6], [46], [31], [7], [49], [50] AND OUR METHOD ON SHAOSHAN DATASET

Method	Completeness	Correctness	Quality
Zang et al. [6]	0.7786	0.7135	0.5963
Unet [46]	0.5301	0.5857	0.3855
Original ResNet[31]	**0.8836**	0.4566	0.4307
ResidualUnet[7]	0.7454	0.9149	0.6970
PSPNet[49]	0.6888	**0.9434**	0.6615
ESPNet[50]	0.7431	**0.8882**	0.6795
Ours	0.8247	0.8443	**0.7159**

TABLE IV
COMPARISON RESULTS OF THREE DIFFERENT SCANNING SIZES (20 × 20, 30 × 30, 40 × 40) RESNET-50 AND OUR MULTISCALE-HIGH-ORDER RESNET-50 ON SHAOSHAN DATASET

method	completeness	correctness	quality
LDMM+ResNet-50 (20×20)	0.7250	0.6343	0.5112
LDMM+ResNet-50 (30×30)	**0.8570**	0.6756	0.6072
LDMM+ResNet-50 (40×40)	0.8528	0.7351	0.6523
LDMM+ResNet-50 (multiscale-high-order)	0.8247	**0.8443**	**0.7159**

where TP, FN and FP denote true positive, false negative and false positive, respectively. As seen from table III, our method achieves best results among seven comparing methods in quality, which proves the good performance of our method. The completeness of our method is also only slightly lower than the best result. Although U-net and original ResNet perform well in many areas, they obtain lower correctness and quality compared with our method on Shaoshan dataset. The reason is probably due to the complexity of background in remote sensing images. Benefiting from the initial segmentation of LDMM, our method can remove parts of background areas first. Besides, the mutiscale-high-order strategy further improves the performance of original ResNet. Thus, our approach obtains better results. Another reason why we achieve a better performance than original ResNet is the pre-segmentation by LDMM, which removes most backgrounds having different gray-value distribution with roads. Residual Unet, PSPNet and ESPNet obtain higher correctness than our method. However, they achieve much lower completeness, as well as lower quality, than our method.

3) Effectiveness of Multiscal-High-Order Strategy: To further illustrate the effectiveness of multiscale-high-order, we compare the performance of our method with and without multiscale-high-order strategy. Table IV shows the comparison results of three different scanning sizes (20 × 20, 30 × 30 and 40 × 40) ResNet-50 and our multiscale-high-order ResNet-50 on Shaoshan dataset. Obviously, our proposed mutiscale-high-order strategy obtains better results than a single scanning size ResNet-50 model. Compared with three ResNet-50 models using scanning sizes of 20 × 20, 30 × 30, 40 × 40, we

[1] https://github.com/Vladkryvoruchko/PSPNet-Keras-tensorflow,
[1] https://github.com/DuFanXin/deep_residual_unet,
[1] https://github.com/sacmehta/ESPNet

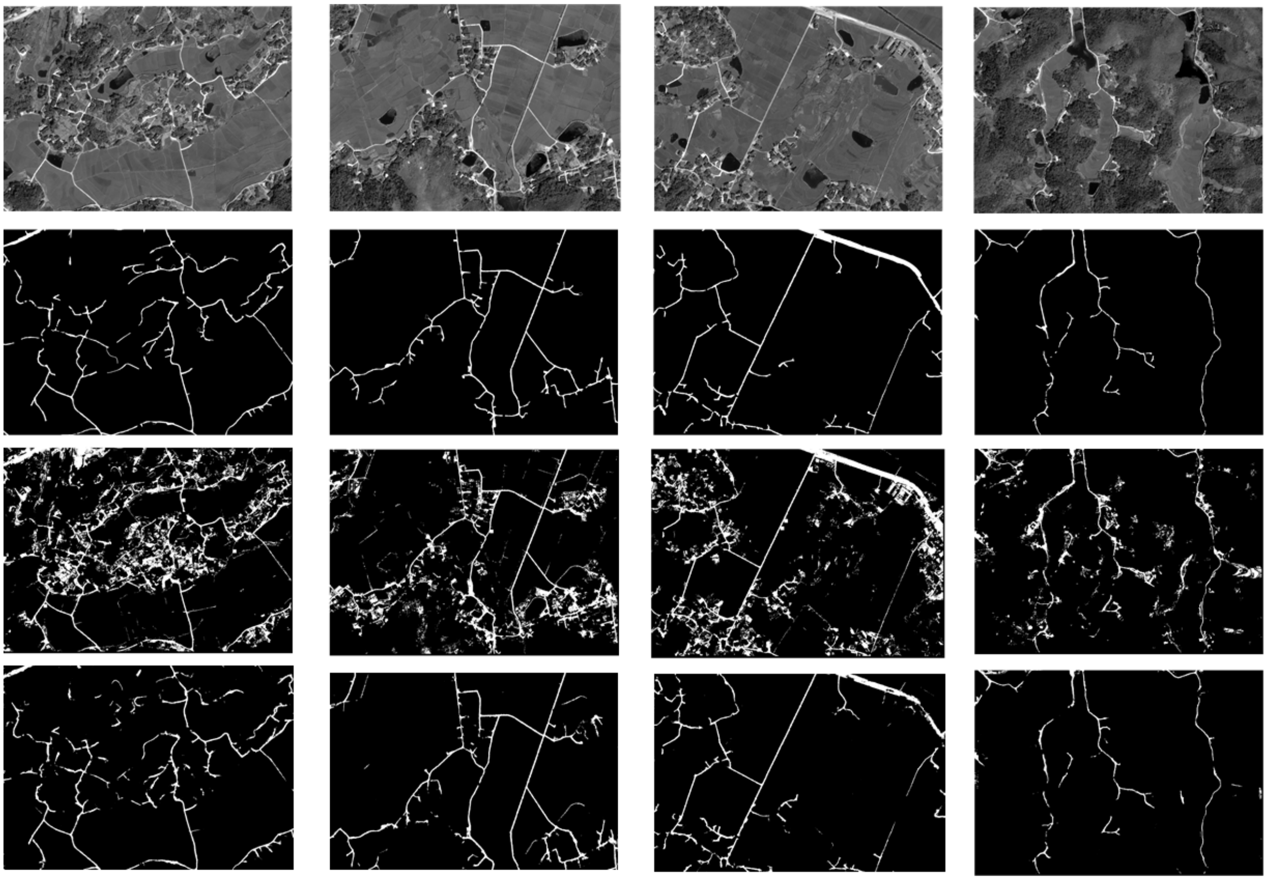


Fig. 8. Several test results of our method on Shaoshan dataset. The rows 1-4 are orginal images, labes, results of LDMM, final results of our method, respectively.

TABLE V

PROCESSING TIME COMPARISON OF DIFFERENT 7 METHODS FOR EACH 1131 × 1589 TEST IMAGE

Method	Implementation Environment	Time
Zang et al. [6]	W-2102|04 2.9GHz CPU	22.51(second)
Unet [46]	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	0.662(second)
Original ResNet[31]	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	9 (hour)
ResidualUnet[7]	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	0.803(second)
PSPNet[49]	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	1.5481(second)
ESPNet[50]	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	**0.2194(second)**
ours	W-2102|04 2.9GHz CPU +NVIDIA 1080ti GPU	2.1(hour)

improve about 21%, 17%, 11% of correctness, respectively. The quality is also improved about 20%, 11%, 6%, respectively. In completeness, multiscale-high-order is only slightly worse than the best result among four strategies.

4) Processing Time Comparison: In the final experiment, we analyze the computational time cost for different methods. Table V shows the processing time comparison of different seven methods for each 1589 × 1131 test image. From table V, we can see ESPNet runs fastest among seven methods. Although our method cost 2.1 hours for per 1589 × 1131 test image, it is an acceptable computational time for obtaining a better performance. The long processing time of CNN model methods is due to the complex neural network convolution operations for each potential positive pixel position. In our following work, we will focus on improving the computational efficiency when using deep CNN model for road extraction from large remote sensing images.

V. CONCLUSION

Road extraction from remote sensing images is still an attractive and challenging task. Gray value distribution and structure feature information are both important for robust road extraction from remote sensing images. In this paper, we proposed a framework for road extraction from remote sensing images, which introducing a local Dirichlet mixture model for pre-segmentation utilizing gray-value distribution information and a multiscale-high-order deep learning strategy for catching high-order structure context information. First, we used the proposed local Dirichlet mixture model to coarsely segment the image into 2 classes (potential road and background). Benefitting from the local strategy, our segmentation is much faster and more accurate compared with original Dirichlet mixture model. Through segmentation of local Dirichlet mixture model, most backgrounds having different gray-value distribution with roads were removed. Then, we trained three different scanning sizes of ResNet-50 models using deep learning. We merged the results of three ResNet-50 models to train a high-order ResNet-50 model to catch structure context information for the final road extraction. Finally, the trained multiscale-high-order ResNet-50 was used to extract road areas based on the segmentation results of local Dirichlet mixture model. Experimental results on Shaoshan dataset illustrated the satisfactory performance of our method compared with other six state-of-the-art methods. We achieved correctness and quality as high as 0.8443 and 0.7159, respectively. And the completeness of our method was only slightly worse than the best result among compared methods. The experiments also proved the effectiveness of multiscale-high-order strategy. Compared with single ResNet model, our multiscale-high-order strategy can greatly improve the correctness and quality while got only slightly worse completeness than the best one. The main shortage of our method is the processing speed, which we will focus on in our following work.

References

[1] 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 Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017.

[2] 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 J. Photogram. Remote Sens.*, vol. 138, pp. 151–163, Apr. 2018.

[3] M. O. Sghaier and R. Lepage, "Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1946–1958, May 2016.

[4] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. D. Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, Aug. 2017.

[5] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 238–248, Jan. 2017.

[6] Y. Zang, C. Wang, Y. Yu, L. Luo, K. Yang, and J. Li, "Joint enhancing filtering for road network extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1511–1525, Sep. 2016.

[7] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2017.

[8] Y. Zang, C. Wang, L. Cao, Y. Yu, and J. Li, "Road network extraction via aperiodic directional structure measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3322–3335, Jun. 2016.

[9] M. Li, A. Stein, W. Bijker, and Q. Zhan, "Region-based urban road extraction from VHR satellite images using binary partition tree," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 44, pp. 217–225, Feb. 2016.

[10] C. Poullis, "Tensor-Cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 95, pp. 93–108, Sep. 2014.

[11] C. Hu, W. Fan, J. Du, and Y. Zeng, "Model-based segmentation of image data using spatially constrained mixture models," *Neurocomputing*, vol. 283, pp. 214–227, Mar. 2017.

[12] W. Fan, N. Bouguila, J.-X. Du, and X. Liu, "Axially symmetric data clustering through Dirichlet process mixture models of Watson distributions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1683–1694, Jun. 2018.

[13] 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 Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1238–1242, Aug. 2017.

[14] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4785–4794, Oct. 2016.

[15] G. Cheng, F. Zhu, S. Xiang, and C. Pan, "Road centerline extraction via semisupervised segmentation and multidirection nonmaximum suppression," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 545–549, Apr. 2016.

[16] 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, pp. 407–420, Sep. 2016.

[17] Z. Hui, Y. Hu, S. Jin, and Y. Z. Yevenyo, "Road centerline extraction from airborne LiDAR point cloud based on hierarchical fusion and optimization," *ISPRS J. Photogramm. Remote Sens.*, vol. 118, pp. 22–36, Aug. 2016.

[18] L. Courtrai and S. Lefèvre, "Morphological path filtering at the region scale for efficient and robust road network extraction from satellite imagery," *Pattern Recognit. Lett.*, vol. 83, pp. 195–204, Nov. 2016.

[19] 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, pp. 88–95, Nov. 2016.

[20] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3359–3372, Jun. 2014.

[21] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1538–1544, Oct. 2012.

[22] W. Shi, Z. Miao, Q. Wang, and H. Zhang, "Spectral–spatial classification and shape features for urban road centerline extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 788–792, Apr. 2014.

[23] L. Wei, P. Xing, J. Zeng, J. Chen, R. Su, and F. Guo, "Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier," *Artif. Intell. Med.*, vol. 83, pp. 67–74, Nov. 2017.

[24] L. Wei, S. Wan, J. Guo, and K. K. L. Wong, "A novel hierarchical selective ensemble classifier with bioinformatics application," *Artif. Intell. Med.*, vol. 83, pp. 82–90, Nov. 2017.

[25] Y. Chen *et al.*, "Fast density peak clustering for large scale data based on kNN," *Knowl.-Based Syst.*, to be published.

[26] Y. Chen, S. Tang, N. Bouguila, C. Wang, J. Du, and H. Li, "A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data," *Pattern Recognit.*, vol. 83, pp. 375–387, Nov. 2018.

[27] S. Movaghati, A. Moghaddamjoo, and A. Tavakoli, "Road extraction from satellite images using particle filtering and extended Kalman filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2807–2817, Jul. 2010.

[28] S. Leninisha and K. Vani, "Water flow based geometric active deformable model for road network," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 140–147, Apr. 2015.

[29] I. Grinias, C. Panagiotakis, and G. Tziritas, "MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images," *ISPRS J. Photogram. Remote Sens.*, vol. 122, pp. 145–166, Dec. 2016.

[30] A. Troya-Galvis, P. Gançarski, and L. Berti-Équille, "Remote sensing image analysis by aggregation of segmentation-classification collaborative agents," *Pattern Recognit.*, vol. 73, pp. 259–274, Jan. 2017.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Pssrocess. Syst. (NIPS)*, 2012, pp. 1097–1105.

[33] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[34] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[36] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, Apr. 2015, pp. 1–14.

[38] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[39] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[40] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[41] B. Zhong, B. Bai, J. Li, Y. Zhang, and Y. Fu, "Hierarchical tracking by reinforcement learning-based searching and coarse-to-fine Verifying," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2331–2341, May 2019.

[42] Q. Zhou, B. Zhong, Y. Zhang, J. Li, and Y. Fu, "Deep alignment network based multi-person tracking with occlusion and motion reasoning," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1183–1194, May 2018.

[43] X. Liu, J. Geng, H. Ling, and Y.-M. Cheung, "Attention guided deep audio-face fusion for efficient speaker naming," *Pattern Recognit.*, vol. 88, pp. 557–568, Apr. 2019.

[44] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[45] W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 155–165, Mar. 2016.

[46] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[47] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1633–1644, May 2018.

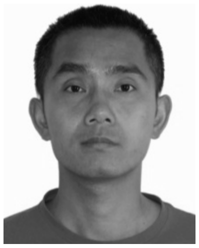
[48] Y. Zhuang, F. Yang, L. Tao, C. Ma, and W. Gao, "Dense relation network: Learning consistent and context-aware representation for semantic image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3698–3702.

[49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

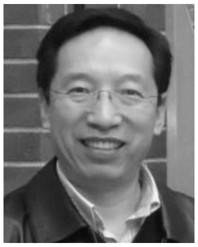
[50] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 552–568.

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

Wentao Fan received the M.Sc. and Ph.D. degrees in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2009 and 2014, respectively. He is currently an Associate Professor with the Department of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include machine learning, computer vision, and pattern recognition.



Bineng Zhong received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2004, 2006, and 2010, respectively. From 2007 to 2008, he was a Research Fellow with the Institute of Automation and Institute of Computing Technology, Chinese Academy of Sciences. From September 2017 to September 2018, he was a Visiting Scholar with Northeastern University, Boston, MA, USA. He is currently a Professor with the School of Computer Science and Technology, Huaqiao University, Xiamen, China. His current research interest includes computer vision.



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 with the School of Information Science and Engineering, Xiamen University, China. His current research interests include information extraction from earth observation images and 3-D surface reconstruction from mobile laser scanning point clouds.

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 from 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.

Cheng Wang (SM'16) received the Ph.D. degree in information communication engineering from the National University of Defense Technology, China, in 2002. He is currently a Professor with 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.