AUTOMATIC EXTRACTION OF ROAD CENTERLINES AND EDGE LINES FROM AERIAL IMAGES VIA CNN-BASED REGRESSION

Yujun Wei[1], Xiangyun Hu[1]*, Mi Zhang[1], Yue Xu[1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, China
(e-mail address: weiyujun@whu.edu.cn; huxy@whu.edu.cn; mizhang@whu.edu.cn; yuexu41@Whu.edu.cn)

Commission II, WG II/III

KEY WORDS: Road extraction, Convolution Neural Network, Regression, Confidence map, Road Width

ABSTRACT:

Extracting roads from aerial images is a challenging task in the field of remote sensing. Most approaches formulate road extraction as a segmentation problem and use thinning and edge detection to obtain road centerlines and edge lines, which could produce spurs around the extracted centerlines/edge lines. In this study, a novel regression-based method is proposed to extract road centerlines and edge lines directly from aerial images. The method consists of three major steps. First, an end-to-end regression network based on CNN is trained to predict confidence maps for road centerlines and estimate road width. Then, after the CNN predicts the confidence map, non-maximum suppression and road tracking are applied to extract accurate road centerlines and construct road topology. Meanwhile, Road edge lines are generated based on the road width estimated by the CNN. Finally, in order to improve the connectivity of extracted road network, tensor voting is applied to detect road intersections and the detected intersections are used as guidance for the overcome of discontinuities. The experiments conducted on the SpaceNet and DeepGlobe datasets show that our approach achieves better performance than other methods.

1. INTRODUCTION

Road extraction from high-resolution remote sensing images is an essential task in the field of remote sensing. It has a wide range of applications, such as vehicle navigation, urban planning, autonomous driving and automatic digital line graphic making. Although various methods have been proposed in recent years, road extraction is still a challenging task because of the considerable variation in road shape, color and context characteristics. Besides, roads is often occluded by objects such as shadows and trees, thereby increasing the difficulty of extraction.

To deal with road extraction task, many CNN-based methods (Panboonyuen T et al., 2017) have been proposed. However, most of these approaches formulated road extraction as a segmentation problem. In automatic map construction, the road centerlines and edge lines are needed. Thus, the thinning operation and edge detection are followed to attain the road centerlines and edge lines, the segmentation-based methods have several defects: (1) Thinning operation could easily bring about spurs around the extracted centerlines. (2) The road topology is not taken into account. So it is of great practical significance with a framework to directly extract road centrelines and edge lines from satellite images. In human pose estimation, researchers aim to locate the anatomical keypoints. The basic idea of pose estimation methods is to produce 2D belief maps for the location of each part. The belief maps encode the spatial uncertainty of each keypoint's location.

Inspired by the belief maps in human pose estimation, for overcoming the above-mentioned shortcomings in the existing methods, this study proposes an end-to-end regression network to learn confidence maps for road centerlines and road width map. A confidence map represents the probability that each pixel is lying on the road centerlines and a road width map indicates the width of road where each pixel is located. After the regression network predicts the confidence map and road width map, a simple non-maximum suppression (NMS) method and road tracking are applied to obtain accurate road centerlines. Road edge lines are then extracted based on the tracked centerlines and estimated road width. Finally, in order to improve the connectivity of extracted road network, tensor voting is applied to detect road intersections and we use detected intersections as the guidance for the overcome of the discontinuities. The final output of our algorithm is file of road centerlines and edge lines with shp format, which can be directly used in automatic road map generation.

In this study, we use a multi-task learning strategy to jointly learn confidence map and road width map, which could not only improve the efficiency of computation but also enhance the generalization ability of the network. Considering roads have long continuous shape structure, thus the spatial relationship is essential for road extraction. The astrous convolutions (Chen et al., 2018) with different rates are proven to effectively capture the context information, which are also adopted in our network.

This study conducts experiments on the DeepGlobe (Demir et al., 2018) and SpaceNet (Etten et al.,2018) datasets, then we compare our approach to other road extraction methods. Our method achieves equal performance to the state of the arts.

Below the related works of road extraction is presented in Section 2. Methods for road centerlines extraction, edge lines generation and overcome of discontinuities are explained in Section 3. Experiment procedure and results are shown in Section 4. Conclusions are presented in Section5.

*Corresponding author

2. RELATED WORK

In recent years, there are many works attempting to extract roads from aerial images. Most of these approaches can be divided into two classes: methods based on heuristic knowledge or methods based on machine learning. The heuristic methods generally utilize some prior knowledge in road extraction, such as edges, radiometry, texture, geometry, etc. For instance, Hu et al., (2005) proposed a model which describes the radiometry and geometry characteristics of roads. He analysed the profiles along the direction perpendicular to the road direction and extracted ribbon roads form satellite images. Mohammadzadeh et al., (2006) extracted main road networks from IKONOS imagery using an approach based on mathematical morphology and fuzzy logic. Movaghati et al., (2010) made a combination between Extended Kalman filter and a special particle filter (PF) to maintain the robustness of road tracing in region where roads are occluded by obstacles. The road tracer finds and follows different road directions after it reaches a road junction. Shao and Guo et al., (2011) presented an effective and fast approach to detect ribbon-like curvilinear structure from remote sensing images. The key content of the algorithm is a simple assumption: the grey value of center pixel and its near neighbor pixels in road region are lighter than that of pixels in the surrounding region. In contrast, machine learning methods take advantage of the huge data to train models for road extraction. Maurya et al., (2011) used the K-Means clustering to classify each pixel in aerial images into two classes, road and non road. Then the non road area is removed based on the morphological features. Huang et al., (2009) used object-oriented algorithm to extract structural features such as Shape Index and Density, then adopted support vector machines (SVM) to classify regions into road or non road based on multiscale spectral-structural features. Wegner et al., (2013) proposed a higher-order CRF for road labeling, in which the spatial properties of road network are exploited and is represented by the higher-order cliques as the prior for road extraction. Mattyus et al., (2015) made use of Markov random field to inference the location of road centerlines and road width based on the OpenStreetMap (OSM). The algorithm is very efficient and the OSM roads of the world could be segmented in one day.

Recently, the convolutional neural network(CNN) has achieved huge success in computer vision and remote sensing image processing, such as image classification, object detection and semantic segmentation. CNNs have also been used for road extraction from aerial images. Mnih and Hinton et al., (2010) made the first attempt of applying deep learning in the field of road extraction. They adopted restricted Boltzmann machines (RBMs) for urban network extraction. To fully use the context information, a large patch is trained to predict the road map in the center area of it and PCA is adopted for decreasing the dimension of the input image. To get better results, Saito et al., (2016) proposed a CNN model for semantic segmentation of aerial images, the image is segmented into three classes (road, building, background). Besides, a new channel-wise inhibited softmax (CIS) loss function is designed to obtain better segmentation results. Mttyus et al., (2017) taked advantage of CNN models to have an initial road segmentation of aerial imagery and then designed an algorithm to overcome the missing connections in the extracted road topology. Cheng et al., (2017) proposed CasNet, a cascaded convolutional neural (CNN) network to simultaneously conduct road segmentation and road centerline extraction tasks from aerial imagery. However, the distribution between centerlines and background is heavily biased and their method still needs thinning operation, which cannot extract accurate road centerlines and infer the road topology. Inspired by the deep residual learning and U-Net, Zhang et al., (2017) proposed the deep residual U-Net, which combined the deep residual learning with U-Net architecture and is more powerful in road segmentation. Wei et al., (2017) proposed RSRCNN, a CNN model for obtaining refined segmentation of road structures from aerial images, a loss function considering spatial correlation and geometric information of road structure is designed to train the CNN model. Zhou et al., (2018) proposed D-LinkNet for road extraction, the backbone of D-LinkNet is LinkNet and several dilated convolution layers are added in its center part. D-LinkNet won the first place in the DeepGlobe2018 Road Extraction Challenge.

Several work represent road network as an undirected graph. Bastani et al., (2018) proposed roadtracer, a method that used a CNN-based decision function to guide an iterative search process to generate a road graph. Ventura et al., (2018) designed a CNN that predicts the connectivity between the current road nodes and other nodes in its neighbourhood. In polymapper (Li et al., 2019), the author defines the road as closed graph and used Polygon-RNN to detect the position of graph nodes.

3. METHODOLOGY

The workflow of the proposed method is shown in Figure 1. Aerial images are inputs in the study, a regression network is trained to predict the confidence map for road centerline and road width map. At the inference stage, after the network predicts the confidence map and width map, NMS and road tracking are applied to attain road centerlines. Then road edge lines are generated based on the extracted centerlines and predicted road width map. Finally, in order to improve the completeness of extracted road network, tensor voting is applied to detect road intersections and the detected intersections are used to guide the overcome of discontinuities.

3.1 Confidence map for centerlines

The training set is denoted as $S = \{X_n, Y_n, Z_n, n = 1, ..., N\}$, in which $X_n = \{x_j^{(n)}, j = 1, ..., |X_n|\}$ denotes the input aerial image, $Y_n = \{y_j^{(n)}, j = 1, ..., |Y_n|\}$ denotes the corresponding confidence map for road centerlines. $Z_n = \{z_j^{(n)}, j = 1, ..., |Z_n|\}$ denotes the corresponding road width map.

The value of pixel on confidence map represents its probability of lying on the road centerlines, which is shown in Figure 2. The confidence map has the following properties: $y_j^{(n)}$ is local maximum when $x_j^{(n)}$ is on the centerlines and the value of $y_j^{(n)}$ gradually decreases as the distance between $x_j^{(n)}$ and road centerlines becomes larger. The confidence map Y_n is defined as follows:

$$y_j^{(n)} = e^{-D_C(x_j^{(n)})^2/2\sigma^2} \tag{1}$$

$$D_C(x_j^{(n)}) = \min_{x_k in\ centerlines} \left\| x_j^{(n)} - x_k \right\| \tag{2}$$

where $D_C(x_j^{(n)})$ denotes the minimum distance from the j-th pixel $x_j^{(n)}$ to pixel x_k on the road centerlines. σ controls the spread of the peak. In our approach, we set $\sigma = 5$. In most datasets for road extraction, the road label is a binary mask, in which the values of the pixels in road areas are 1 and the values

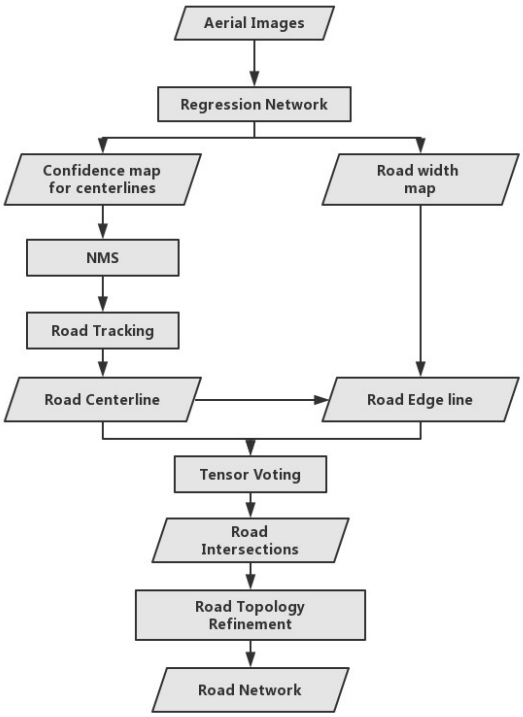
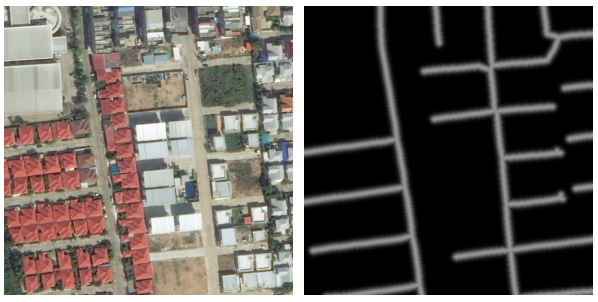


Figure 1. Workflow of the proposed method

of the pixels belonged to background are 0. Thus, we need to convert the road mask to the confidence map for road centerlines. We adopt a thinning algorithm to obtain road centerlines from road mask and then generate the road confidence maps. The confidence map for centerlines is calculated as Eq. (1).



(a) (b)

Figure 2. Confidence map for road centerlines. (a) Input image. (b) Corresponding confidence map.

3.2 Road width map

In order to estimate the road width of each pixel on extracted road centerlines, the study proposes road width map which indicates the width of road where each pixel is located. A regression network is trained to predict road width map for input images. The ground-truth for road width map is generated based on road binary mask either. For each pixel on the roads, the width of which is calculated along the direction perpendicular to road direction, as shown in Figure 3. The road width map Z_n is calculated as follows :

$$Z_j^{(n)} = \begin{cases} W(x_j^{(n)}) & if\ P(x_j^{(n)}) = 1 \\ 0 & ifP(x_j^{(n)}) = 0 \end{cases} \quad (3)$$

where $W(x_j^{(n)})$ denotes the road width of the j-th pixel $x_j^{(n)}$. $P(x_j^{(n)})$ denotes whether $x_j^{(n)}$ is in the road region. $P(x_j^{(n)}) = 1$ denotes that $x_j^{(n)}$ is in the road region and $P(x_j^{(n)}) = 0$ denotes that $x_j^{(n)}$ is in the background.

Figure 3. For pixel p and q, the width of which ($w1$, $w2$) is calculated along the direction perpendicular to road direction.

3.3 Network architecture

The whole network is shown in Figure 4. We adopt encoder-decoder architecture to simultaneously predicts confidence map and width map. We choose the pretrained ResNet (He et al, 2016) as the encoder. The network is split into two branches, the first decoder predicts the confidence map while the second predicts the width map. The two decoders share the same architecture. Each decoder has five upsampling layers to gradually upsample the feature map by a factor of 2. Each upsampling layer is followed by 2 convolution layers to generate dense predictions.

Road generally has a long continuous shape structure, thus context information and spatial relationship are important in road recognition. Dilated convolutions with different dilate rates can effectively increase the receptive field of network while preserving the details, so this study adds additional dilated convolution layers in the center part of the network.

The confidence map and width map predicted by our network are denoted as $\widetilde{Y}$ and $\widetilde{Z}$. The ground-truth for confidence map and width map are denoted as Y and Z. In this study, mean-squared loss is adopted as the loss function of our network. The loss for predicted confidence map is denoted as L_Y while the loss for predicted width map is denoted as L_Z. L_Y and L_Z are calculated as following.

$$L_Y = \frac{1}{N}\sum_p (\widetilde{Y}(p) - Y(p))^2 \quad (4)$$

$$L_Z = \frac{1}{N}\sum_p (\widetilde{Z}(p) - Z(p))^2 \quad (5)$$

where p denotes position of pixel in the confidence and width maps. N denotes the number of pixels in the map. The total loss function of the network is the sum of L_Z and L_Y, which is defined as

$$Loss = L_Y + L_Z \tag{6}$$

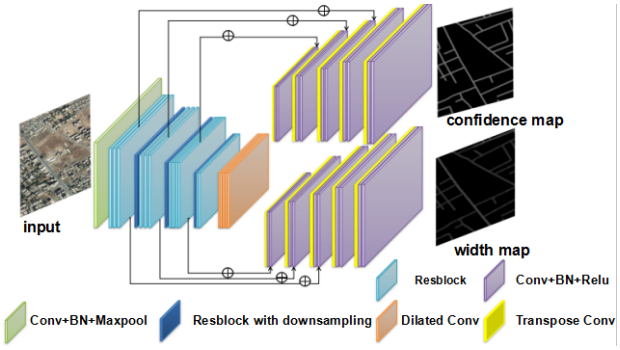


Figure 4. Our architecture is split into two branches: top and bottom, which simultaneously predicts confidence and road width maps.

3.4 Extraction of centerlines and edge lines

Since the values of pixels on centerlines in the confidence map are local maximums along the direction perpendicular to direction of road. After the network predicts the confidence map for centerlines and road width map, a Canny-like non-maximum (NMS) is applied to the confidence map to obtain accurate centerlines. Given the confidence map M predicted by CNN, the direction θ perpendicular to the road direction at position $p(x, y)$ is calculated as follows.

$$\theta = \tan^{-1}(Dy, Dx) \tag{7}$$

where Dx and Dy denote the Horizontal and Vertical gradients. If the value of $p(x, y)$ on M is local maximum along direction θ, p is on centerlines.

Although the centerlines are extracted after NMS, the results are simply binary images (as shown in Figure 5), which lack the road topology. Thus in order to construct road topology, we take advantage of road tracking to track road centerlines.

To track road centerlines, firstly, a point on centerlines is selected as the start point, and the direction of the road at the start point is calculated. The binary image for road centerlines is denoted as C, $C = \{c_{i,j}|i = 1, ...H, j = 1..., W\}$ where $c_{i,j} = 1$ for pixel (i, j) on the centerlines and $c_{i,j} = 0$ for pixel (i, j) on the background. Given the direction $\theta_{current}$ of current trace point $(x_{current}, y_{current})$, the positions of candidates for the next trace point are calculated as follows.

$$\begin{bmatrix} x_{s,t} \\ y_{s,t} \end{bmatrix} = \begin{bmatrix} x_{current} + \cos(\theta_{current} + t) \times S \\ y_{current} + \sin(\theta_{current} + t) \times S \end{bmatrix} \tag{8}$$

where t is the change of road direction. S is the suitable size of the distance between current trace point and the next point. In this study, $t \in (0, \pm 1, ..., \pm 10)$, $S = 15$.

The next trace point (x_{next}, y_{next}) is calculated using the following formula.

$$(x_{next}, y_{next}) = (x_{s,t_{min}}, y_{s,t_{min}}) \tag{9}$$

$$t = \underset{t}{argmin} C(x_{s,t}, y_{s,t}) = 1 \tag{10}$$

The road direction θ for the next trace point is updated as follows.

$$\theta_{current} = \theta_{current} + t \tag{11}$$

After road centerlines are tracked, road edge lines are generated based on the tracked centerlines and road width map predicted by the CNN model. Let W denotes the predicted road width map, p denotes pixel on the tracked centerlines. The locations of pixels on road edge lines are calculated as following.

$$\begin{cases} x = x_p \pm W(p) \times (-\sin\theta_p) \\ y = y_p \pm W(p) \times \cos\theta_p \end{cases} \tag{12}$$

where (x_p, y_p) denotes position of p and θ_p denotes road direction at p. The extracted road centerlines and edge lines are shown in Figure 5.

(a) (b) (c) (d)

Figure 5. Extraction of centerlines and edge lines. (a) Input image. (b) Binary map for centerlines. (c) Tracked road centerlines. (d) Extracted road edge lines.

3.5 Refining road topology

After the previous road tracking, the main road network has been extracted. However, there are still some gaps and isolated road fragments in the extracted road network. Most of isolated segments should be connected to other roads to generate intersections, as shown in Figure 6. Therefore, this study used tensor voting algorithm (Maggiori et al., 2015) to overcome the discontinuities. Though intersections could be directly predicted by CNN, tensor voting is more simple and doesn't need training, which is a more generic for road network refinement.

Tensor voting is a robust method for perceptual grouping. It firstly encodes input space points as stick-shaped tensors or ball-shaped tensors. After encoding the input points into perfect tensors (Maggiori et al., 2015), the information they encode is propagated to their neighbourhood in the voting procedure. After the first voting, the tensors for input points are refined and a second voting is carried out.

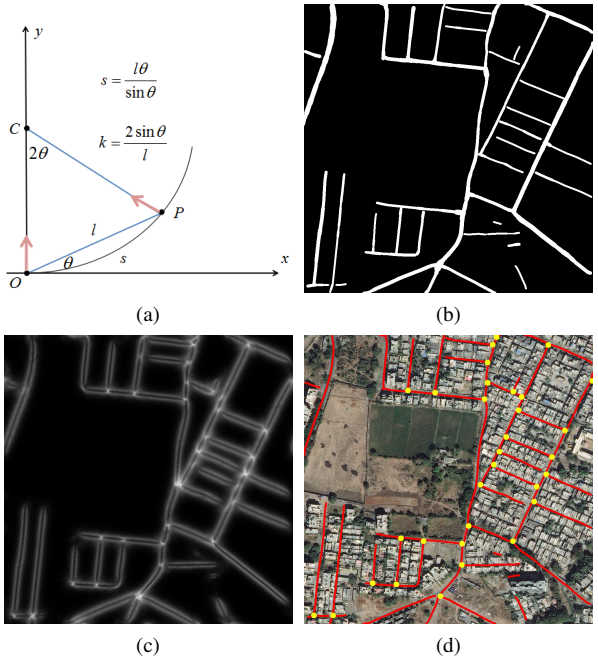


Figure 6. Intersections detected by tensor voting. (a) tensor voting. (b) Binary map for road region. (c) Saliency of ball tensors. (d) Detected intersections are shown in yellow.

In the tensor encoding procedure, for each pixel p in the road region, the normal vector of p is $n = (n_x, n_y)$. p is encoded as tensor $T = \begin{bmatrix} n_x n_x & n_x n_y \\ n_x n_y & n_y n_y \end{bmatrix}$. T can be decomposed as follows.

$$T = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T = (\lambda_1 - \lambda_2) e_1 e_1^T + \lambda_2 (e_1 e_1^T + e_2 e_2^T) \tag{13}$$

where λ_1 and λ_2 are eigenvalues of T and $\lambda_1 \geq \lambda_2$, e_1, e_2 are eigenvectors. $\lambda_1 - \lambda_2$ is the saliency for stick tensor and λ_2 is the saliency for ball tensor.

Voting is carried out after points in road region have been encoded. Tensors propagate their information to other points in the neighbourhood, as shown in Figure 6 (a). Assuming that P is the voting point and O is the receiver. The salience of vote from P received by O is calculated as follows.

$$DF(\sigma) = e^{-(s^2 + c\kappa^2/\varepsilon^2)} \tag{14}$$

where s denotes the length of arc along the osculating circle from P to O and κ denotes the curvature. ε denotes the scale parameter. The decay of salience is controlled by c. The vote of P received by O is calculated as follows.

$$SV(T, v) = \begin{cases} DF(v) R_{2\theta} T R_{2\theta}^T, & if \ -\pi/4 \leq \theta \leq \pi/4 \\ 0, & otherwise \end{cases} \tag{15}$$

where θ is the angle subtended by the arc of the osculating circle from P to O. $R_{2\theta}$ is the rotation matrix for 2θ. T is the tensors for P.

After tensor voting, the ball saliency of intersections is higher than other points in its neighbourhood, which is shown in Figure 6(c). The intersections are extracted after applying NMS to the saliency map of ball tensors. The detected intersections are shown in Figure 6(d).

Figure 7. Intersections guide overcome of discontinuities. (a) inital road network. (b) refined road network.

The detected intersections are used as the guidance for the overcome of discontinuities. If the intersections are on the extended lines of road line segments, the road fragments are connected to the corresponding intersections. Figure 7 shows the refined road topology.

4. EXPERIMENTS

4.1 Dataset

This study conducts experiments on DeepGlobe and SpaceNet datasets. DeepGlobe dataset consists of 6226 aerial images. The satellite imagery used in DeepGlobe is sampled from the DigitalGlobe+Vivid Image dataset, the spatial resolution of that is $1m^2/pixel$. We randomly select 4626 images for the training part and 1600 for the testing part. SpaceNet dataset consists of 3347 images, the ground resolution of which is 30 cm/pixel. This dataset includes four areas: Las Vegas, Paris, Shanghai, and Khartoum. We split the dataset into 2780 images for training and 567 for testing.

4.2 Implementation details

We implement the proposed network using the Pytorch framework. Encoder is initialized using the pretrained model on ImageNet dataset. The network is optimized using RMSprop with learning rate policy of poly. The hyper parameters of our model include initial learning rate ($2e^{-4}$), mini-batch size (2) and max epoches(300).

4.3 Evaluation Metrics

Two different measures are used to evaluate the quality of extracted road networks: a classical measure and a measure named connectivity to evaluate the connectivity of topology.

The classical measure (Heipke et at., 1997) consists of recall, precision and F1-score. Their definitions are presented as follows.

$$recall = \frac{n_m^*}{n_t^*} \tag{16}$$

$$precision = \frac{n_m}{n_t} \tag{17}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{18}$$

where n_m^* denotes the length of reference road path in the buffer of extracted road network, n_t^* denotes the length of reference road network. Similarly, n_m denotes the length of extracted

first row: DLinkNet **middle row: ours** **bottom row: gt**

Figure 8. Examples of road extraction results on the DeepGlobe and SpaceNet datasets. The first row shows road centerlines and edge lines extracted by DLinkNet, the middle row shows results of the proposed method and the bottom row shows the ground-truth. We overlayed the extracted road centerlines(yellow) and edge lines(red) on the image. Our approach extracts more complete road network while has lower error rate than other methods.

road path in the buffer of reference road network and n_t denotes the length of extracted road network. The buffer width is set as 3 pixels in the experiments.

The Connectivity C (Ventura et al., 2018) is defined as the ratio of continuous road segments, the definition of which is presented as follows.

$$connectivity = \frac{N_{con}}{N_{gt}} \tag{19}$$

where N_{con} denotes the number of continuous road segments and N_{gt} denotes the number of gt segments.

4.4 Quantitative Results

Our approach has been compared with some deep learning-based road extraction methods, Unet and DLinkNet. This study make experiments on the test set of the DeepGlobe and SpaceNet datasets. U-Net structure is widely used in biomedical images segmentation and has shown great performance in road segmentation. D-LinkNet performed well in the DeepGlobe challenge and won the first place in the DeepGlobe Road Extraction Challenge. We report performance in terms of mean recall, precision, F1-score and connectivity across the two datasets.

On the DeepGlobe dataset, our method outperforms other methods in terms of mean recall, precision and F1-score (as shown in Table I). Specifically, in comparison with D-LinkNet, our method obtains increments of 0.69% in mean recall and 0.35% in mean precision, thereby indicates that ours method can extract more complete road networks while has a lower error rate than other methods. Furthermore this study calculates F1-scores to assess the overall performance of extracted road topology. Our approach obtains increments of 0.54% in mean F1-score. In addition, the proposed method obtains increments of 0.52% in mean connectivity. The quantitative results show that our method remarkably surpasses other methods in extracting high-quality road networks. This study then conducts experiments on the SpaceNet dataset to further evaluate the performance of proposed method. Results are shown in Table II. The proposed method obtains increments of 1.48% and 0.53% in mean reall and precision, obtains increments of 1.05% and 2% in F1-score and connectivity with respect to D-LinkNet. The reuslts indicate that the proposed method still achieves higher performance on SpaceNet dataset.

4.5 Qualitative Results

Figure 8 shows some predicted results of the methods mentioned above on the test set of the SpaceNet and DeepGlobe

Method	Recall	Precision	F1-score	Connectivity
U-Net	0.7868	0.8174	0.8018	0.7663
D-LinkNet	0.8170	0.8756	0.8452	0.8041
ours	**0.8239**	**0.8791**	**0.8506**	**0.8093**

Table 1. Performance on the DeepGlobe dataset. Recall and precision are calculated using a distance threshold of 3 (in pixels)

Method	Recall	Precision	F1-score	Connectivity
U-Net	0.5733	0.6157	0.5937	0.6132
D-LinkNet	0.6015	0.6669	0.6325	0.6486
ours	**0.6163**	**0.6722**	**0.6430**	**0.6686**

Table 2. Performance on the SpaceNet dataset. Recall and precision are calculated using a distance threshold of 3 (in pixels)

datasets. The proposed method extracts more complete road networks and the extracted road networks have less discontinuities, especially in urban area where roads are often occluded by buildings and shadows. Although the proposed method can extract relatively complete road networks, it still has a lower error rate and does not produce more incorrect road fragments. The extracted road edge lines shown in Figure 8 indicate that the road width estimated by CNN is relatively accurate compared with the ground-truth.

Generally, in DeepGlobe and SpaceNet datasets, the proposed method achieves higher performance in road topology extraction against the baselines (Unet, DLinkNet). However, there are still some false road fragments and discontinuities in the extracted road neworks, especially in dense urban areas where roads are frequently occluded and have visual similarity with some buildings. Thus, it is still a great challenge to extract road networks in dense urban areas.

5. CONCLUSIONS

This study proposes a regression-based method for automatic extraction of road centerlines and edge lines from aerial images. The first step is to train a regression network for predicting confidence maps for road centerlines and road width map. Second, after the CNN predicts the confidence map, NMS and road tracking are followed to attain accurate road centerlines. Road edge lines are generated based on extracted centerlines and road width estimated by the network. Finally, in order to improve the connectivity of extracted road network, tensor voting is applied to detect road intersections and the detected intersections are used as the guidance for the overcome of discontinuities. The major contribution of this study is the introduction of the method that uses a new strategy, which is different from image segmentation to solve the problem of road extraction. We have conduct experiments on the DeepGlobe and SpaceNet datasets and the results indicate that the our approach achieves better performance than some other road extraction methods. In the future we plan to extract road networks in dense urban areas.

REFERENCE

Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S. (2018). Roadtracer: Automatic extraction of road networks from aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4720-4728).

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C. (2017). Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 55(6), 3322-3337.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Raska, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 172-17209). IEEE.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H. (2019). Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3146-3154).

Huang, X., Zhang, L. (2009). Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. International Journal of Remote Sensing, 30(8), 1977-1987.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O. (1997). Evaluation of automatic road extraction. International Archives of Photogrammetry and Remote Sensing, 32(3 SECT 4W2), 151-160.

Li, Z., Wegner, J. D., Lucchi, A. (2019). Topological map extraction from overhead images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1715-1724).

Mohammadzadeh, A., Tavakoli, A., Valadan Zoej, M. J. (2006). Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened ikonos images. The photogrammetric record, 21(113), 44-60.

Maurya, Rohit, P. R. Gupta, Ajay Shankar Shukla. "Road extraction using k-means clustering and morphological operations." 2011 International Conference on Image Information Processing. IEEE, 2011.

Mattyus, G., Wang, S., Fidler, S., Urtasun, R. (2015). Enhancing road maps by parsing aerial images around the world. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1689-1697).

Movaghati, S., Moghaddamjoo, A., Tavakoli, A. (2010). Road extraction from satellite images using particle filtering and extended Kalman filtering. IEEE Transactions on geoscience and remote sensing, 48(7), 2807-2817.

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

Maggiori, E., Manterola, H. L., del Fresno, M. (2014). Perceptual grouping by tensor voting: a comparative survey of recent approaches. IET Computer Vision, 9(2), 259-277.

Máttyus, G., Luo, W., Urtasun, R. (2017). Deeproadmapper: Extracting road topology from aerial images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3438-3446).

Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S. (2017). Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields. Remote Sensing, 9(7), 680.

Shao, Y., Guo, B., Hu, X., Di, L. (2010). Application of a fast linear feature detector to road extraction from remotely sensed imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4(3), 626-631.

Saito, Shunta , T. Yamashita , Y. Aoki. "Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks." Electronic Imaging 60.1(2016):10402-1/10402-9.

Ventura C, Pont-Tuset J, Caelles S. Iterative deep learning for road topology extraction. arXiv preprint arXiv:1808.09814, 2018.

Van Etten, Adam, Dave Lindenbaum, and Todd M. Bacastow. "Spacenet: A remote sensing dataset and challenge series." arXiv preprint arXiv:1807.01232 (2018).

Wegner, Jan D., J. A. Montoya-Zegarra, K. Schindler. "A Higher-Order CRF Model for Road Network Extraction." IEEE Conference on Computer Vision and Pattern Recognition IEEE, 2013.

Wei Y, Wang Z, Xu M. Road Structure Refined CNN for Road Extraction in Aerial Image. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5):709-713.

X. Hu*, C. Tao, 2005. A Reliable and fast ribbon road detector using profile analysis and model-based verification, International Journal of Remote Sensing.

Zhang Z, Liu Q, Wang Y. Road Extraction by Deep Residual U-Net. IEEE Geoscience and Remote Sensing Letters, 2017, PP(99):1-5.

Zhou, Lichen, C. Zhang, and M. Wu. "D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) IEEE, 2018.