Feature extraction and tracking of CNN segmentations for improved road detection from satellite imagery

Tianyu Li, Mary Comer, Josiane Zerubia

FEATURE EXTRACTION AND TRACKING OF CNN SEGMENTATIONS FOR IMPROVED ROAD DETECTION FROM SATELLITE IMAGERY

Tianyu Li, Mary Comer

School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN, USA

Josiane Zerubia

Inria and Université Cote d'Azur
Sophia-Antipolis, 06902, Cedex, France

ABSTRACT

Road detection in high-resolution satellite images is an important and popular research topic in the field of image processing. In this paper, we propose a novel road extraction and tracking method based on road segmentation results from a convolutional network, providing improved road detection. The proposed method incorporates our previously proposed connected-tube marked point process (MPP) model and a post-tracking algorithm. We present experimental results on the Massachusetts roads dataset to show the performance of our method on road detection in remotely-sensed images.

Index Terms— road detection, convolutional network, connected tube model, MPP, road tracking

1 Introduction

Road detection from high-resolution satellite images has been an important research topic for years. Since new roads are constructed frequently, it is essential to update road maps with up-to-date high-resolution satellite images, which makes automated road detection important. However, this problem is very challenging due to the complexity of road networks. For instance, roads can vary greatly in width and color, and there could be noise and occlusions from cars, trees, etc.

Many approaches have been proposed to solve this problem in the past decades, some of them approaching it as a road skeleton detection problem [1], and others as a road segmentation problem [2]. Due to the complexity of road networks, traditional methods can be very complex and ineffective.

With the development of deep learning, many neural networks have been constructed for road segmentation. In [3], Mnih and Hinton trained a supervised deep Neural Network to segment roads in high-resolution aerial images. In [4], Ronneberger et al. proposed the U-Net convolutional network for biomedical image segmentation, which has been shown to work well for road segmentation also. Zhang and Liu et al. [5], built a deep residual U-Net by combining residual unit [6] with U-Net. Buslaev et al. [7] proposed a network consisting of ResNet-34 pre-trained on ImageNet and a de-

coder adapted from vanilla U-Net, which showed superior results in the DeepGlobe 2018 Road Extraction Challenge. Zhou et al. [8] obtained good IoU scores in DeepGlobe 2018 Road Extraction Challenge with their D-LinkNet.

Fig. 1: Upper row: original satellite images; middle row: segmentation results from D-LinkNet34; bottom row: the results of road network extraction by the tube-MPP model.

The road segmentation results from these deep convolutional networks are impressive. These methods can overcome the problem of occlusions and shadows; see Fig 1 (d), (e). However, there still exist some road connectivity problems and incorrect detections, such as in Fig 1 (f). Moreover, the binary segmentation masks are not sufficient for higher level road analysis (such as obtaining certain statistical information

and analyzing the structure of the road network). Filin, Zapara and Panchenko [9] proposed a post-vectoring algorithm combined with their EOSResUNet to extract the roads as vectors. This partially solves the connectivity problem in small gaps but is not able to fill larger gaps or extend roads.

Unsupervised line network extraction by MPP is another choice for road detection. Lacoste, Descombes and Zerubia [10] proposed the Quality Candy model for line network extraction. Schmidt et.al [11] presented a forest point process for the automatic extraction of networks. We proposed in a previous paper [12] a connected-tube MPP model for strip feature extraction. These MPP based models provide good results in many cases, but they assume the gray level variation between a road and the background is large, which is not always true.

Considering the promising road segmentation results from deep convolutional networks, we apply our connected-tube MPP model to the segmentation results. In this way, we accurately extract the roads as connected tubes, which can be used for higher level road analysis. Moreover, we propose a post road-tracking algorithm based on the extracted tube networks for solving road connectivity problems.

The remainder of this paper is organized as follows: In Section 2, we review the segmentation methods based on U-Net and the connected-tube MPP model. In Section 3, we introduce our tracking algorithm. In Section 4, road detection results from the Massachusetts roads dataset [13] are given and discussed. Finally, conclusions are drawn in Section 5.
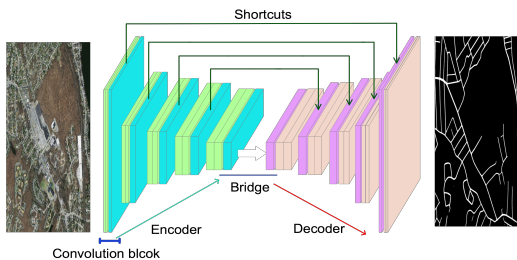

Fig. 2: The architecture of U-Net.

2 Related Work

2.1 Road Segmentation with U-Net

The architecture of U-Net, as in Fig 2, mainly consists of an encoder part, a decoder part and shortcuts. In the encoder and decoder parts, there are 5 basic convolutional blocks each. If we add a residual connection between the input and output for each convolutional block, then we turn the UNet to ResUNet [5]. If we use the ResNet-34 pretrained on ImageNet dataset as the encoder, then the network becomes a LinkNet [14]. Then if we apply the dilation convolution [15] to the bridge part (the middle part of U-Net), the network turns into D-

LinkNet [8].

The training process for U-Net is tricky. First, data augmentation is essential for training the network. The common augmentation strategies include flip, rotate, add noise and change illumination. The selection of loss function is also important. The BCE (binary cross entropy)+ dice coefficient loss [8] works well for road segmentation. And Mosinska [16] proposed a loss function that is aware of the higher-order topological features of linear structures.
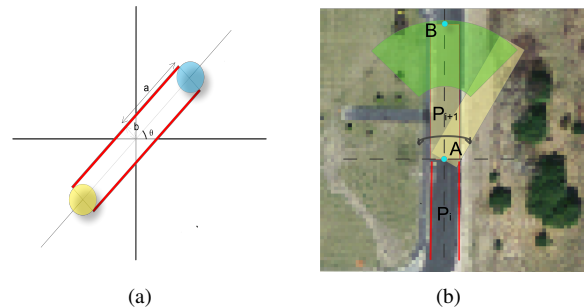

Fig. 3: (a) The shape model of a tube; (b) Tracking illustration.

2.2 Connected-Tube MPP Model

Let Y denote the observed image, S the image lattice, $S \subset \mathbb{R}^2$. A point process on S is a set of points $\{S_1, S_2, ..., S_n\} \subseteq S$, with the random variable S_i representing the random location of the ith point. In a marked point process, for each point S_i, there is an associated mark, which consists of random variables from a mark space M describing an object located at point i. For our tube MPP model, the mark of a tube is defined by the vector (a, b, θ), as in Fig 3 (a). The random variables a, b are the major and minor axes, which control the size of the tube, and θ controls the orientation. The white region is the inner area, the red region is the outer area, which has width of 2 pixels, the blue region and the yellow region are the front connection area and back connection area, respectively. Mark space $M = [a_{min}, a_{max}] \times [b_{min}, b_{max}] \times [0, \pi]$, for some parameters $a_{min}, a_{max}, b_{min}, b_{max}$. A marked object is defined as a vector $W_i = (S_i, M_i) \in W$, where $W \subseteq S \times M$. Let Ω_W be the configuration space, which denotes the space of all possible realizations of W. Then $w = (w_1, ..., w_n) \in \Omega_W$ is a possible object configuration, where n is the number of objects in this configuration. Then the density of the marked point process is given by

$$f(w|y) = \frac{1}{Z} exp\{-V_d(y|w) - V_p(w)\} \qquad (1)$$

where Z is the normalizing constant, $V_d(y|w)$ is the data energy, which describes how well the objects fit the observed image. In this paper we apply the contrast based data energy [17]. $V_p(w)$ is the prior energy introducing the prior knowledge on the object configuration. Here we consider

the overlap prior, tube length prior and connection prior as in [12]. State transitions in the configuration space are realized by three types of kernel: Birth and death kernel, perturbation kernel and switch kernel.

After converting the output (a probabilistic image) of networks from U-Net family to a binary image by a threshold, as in the images in Fig 1 (d), (e), (f), we take this binary image as the input to our connected-tube MPP model. Then we can extract the roads as a connected-tube network, as shown in Fig 1 (g), (h), (i).
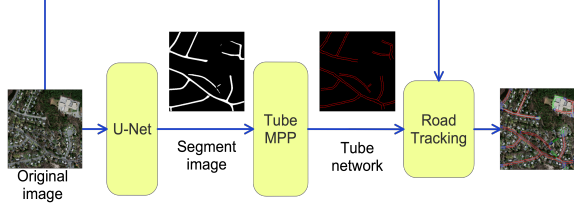

Fig. 4: The framework of the proposed method.

3 Road Tracking

In order to fill some gaps and large missing parts of roads in the segmentation result from deep learning, we propose a tracking algorithm that takes the connected tubes and the original image and extends roads using the method described in this section. The framework is shown in Fig 4.

The open and half open tubes in the extracted connected-tube networks are selected as tracking seeds, which are the tubes with blue edges in Fig 1 (i).

In our work, tracking roads is an iterative process. Given a tube object P_i, our goal is to find a P_{i+1}(as in Fig 3 (b)) which is connected to P_i, has similar features to P_i, and does not conflict with other tube objects. We repeat this process until the whole road has been extracted.

We take the HSV histogram [18] of the inner area of the tube as the color feature C_i of a road segment. This feature is updated in the tracking process by $C_i = \gamma C_{i-1} + (1-\gamma)C$, where γ is the update ratio and C is the color feature of the inner area of the object being tracked. Within the tracking process, we calculate a tracking score TS for each potential tracking object. Tracking score is formulated as:

$$TS = \omega_c dist(C_i, C) - \omega_d dist(C, C_o) + \omega_o ad \quad (2)$$

where C_o is the color feature of the outer area of the tracking object; C_i is the color feature of the road in the ith step; ad is the angle deviation from the original tube object; $\omega_c, \omega_d, \omega_o$ are weights; and $dist(x, y)$ is the Euclidean norm function.

The first term in (2) encourages the tracked object to have similar color to the previous road segments. The second term encourages the tracking object to have larger color difference between inner and outer areas. The third term penalizes the angle deviation.

The details of the tracking process can be seen in Algorithm 1.

Algorithm 1 Road Tracking Function

TrackingFunc(currentObj, colorfeature, TrackingArray):
  if currentObj's front/back end is OPEN **then**
    BestScore = 1000
    A = currentObj's front/back middle point
    for ad in range $(-30, 30)$ **do**
      for len in range $(-2, 2)$ **do**
        newObj.a = currentObj.a+len
        newObj.$\theta = (currentObj.\theta + ang)\%360$
        newObj's back/front middle point = A
        TS_{new} = newObj's score calculated by (2)
        if $TS_{new} <$BestScore **then**
          BestScore = TS_{new}
          BestObj=newObj
        end if
      end for
    end for
    if BestScore$< T_s$ and BestObj not overlap **then**
      TrackingArray.add(BestObj)
      $colorfeature = BestObj.colorfeature * (1-\gamma) + colorfeature * \gamma$
      TrackingFunc(BestObj, colorfeature, TrackingArray)
    end if
  end if

---

4 Experiments

To demonstrate the performance of our method in high-resolution satellite images, we test it on the Massachusetts roads dataset[1] (with the resolution of 1.2 meter). We mainly focus on its ability to extract roads as connected tubes and to solve connectivity problems in road segmentation. For quantitative evaluation, the Jaccard index [19] (ie., Intersection over Union) is used. For sets A and B, it is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3)$$

We apply the D-LinkNet34 [8] for road segmentation and the tube-connected MPP model for road segment extraction. The parameters for the tube model are set as $a_{min} = 4, a_{max} = 16, b_{min} = 2, b_{max} = 8$. For our tracking algorithm, we set $\omega_c = 0.5, \omega_d = 0.3, \omega_o = 0.02, \gamma = 0.88, T_s = 0.38$ by trial and error.

In Fig 5, we present 5 test results to show our method can work for different connection problems. The 512×512 images are cropped from the test set in Massachusetts roads dataset. For the first test, we can see there are some small
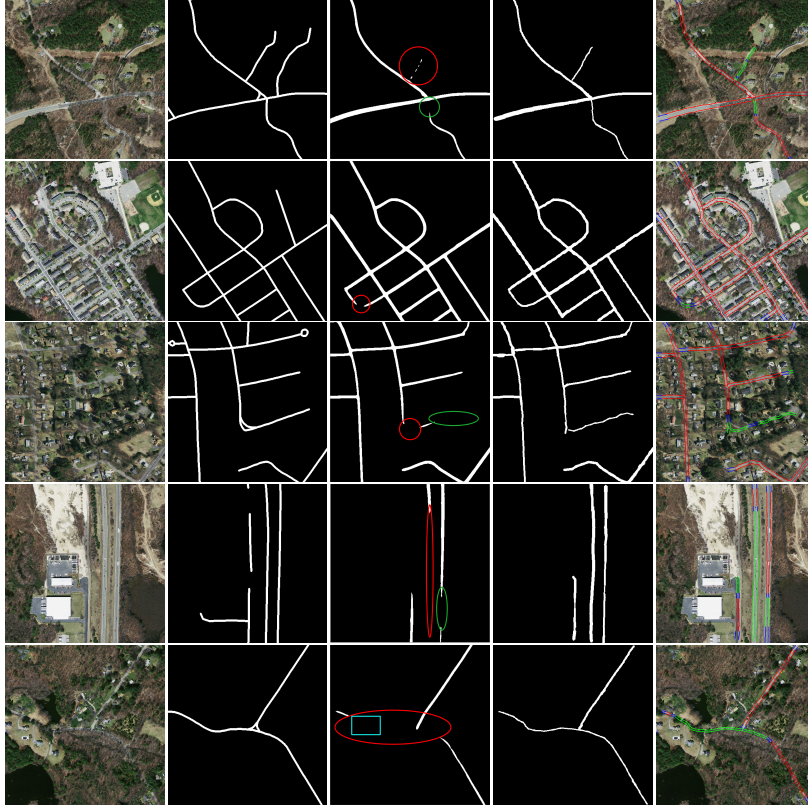

Fig. 5: First column: original images; second column: the ground truth; third column: segmentation results from D-LinkNet34; fourth column: improved segmentation results by our tracking method; fifth column: road network detection results (red tubes are the extracted tubes from connected-tube MPP model; blue tubes are the tracking seeds; green tubes are the tracking tubes).

disconnection parts in the red circle and one segment of road lost in the green circle. In the second test, a small gap appears in the corner of two roads. In the third test, only the middle part of the road has been segmented out. For the fourth test, in the red circle the highway is not segmented out, and in the green ellipse, there is a disconnection problem. By running the proposed algorithm, the gaps have been filled. The fifth test is very challenging, since the missing part of the road in the red ellipse is heavily occluded by trees. With the same experimental parameters, our method would lose the part in the blue box due to heavy occlusion. When we increase the threshold T_s, we detect the whole road as Fig 5 shows. But if we keep this high threshold, there could be some false alarms in other tests. The IoU of these tests are listed in Table 1.

5 Conclusions and Future Work

Road segmentation by deep learning method provide interesting results. However, it still exhibits some connectivity problems and incorrect detection. To deal with the connectivity problems in road segmentation, we propose a post-processing method by combining the tube-connected MPP model and a tracking algorithm. This method not only relieves the prob-lem of disconnection in road segmentation, but also can provide a connected-tube network for higher level road analysis (such as obtaining the statistics of roads or road networks). The experiments on the Massachusetts roads dataset demonstrate the performance of our method. In future research, we will try to learn the parameters in our model automatically and test our method on other applications, such as blood vessel detection.

	test1	test2	test3	test4	test5
DLinkNet34	0.495	0.600	0.607	0.344	0.460
Improved by tracking	0.535	0.635	0.617	0.545	0.569

Table 1: IoU values of the tests in Fig 5.

6 Acknowledgments

7 References

[1] C. Sujatha and D. Selvathi, "Connected component-based technique for automatic extraction of road centerline in high resolution satellite images," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 8, Mar 2015.

[2] R. Alshehhi and P. R. Marpu, "Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 126, pp. 245 – 260, 2017.

[3] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *European Conference on Computer Vision (ECCV)*, Berlin, Heidelberg, 2010, pp. 210–223, Springer.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Cham, 2015, pp. 234–241, Springer.

[5] 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, May 2018.

[6] 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)*, 2016, pp. 770–778.

[7] A. V. Buslaev, S. S. Seferbekov, V. I. Iglovikov, and A. A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 197–1973, 2018.

[8] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 192–1924, 2018.

[9] O. Filin, A. Zapara, and S. Panchenko, "Road detection with eosresunet and post vectorizing algorithm," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 201–2014, 2018.

[10] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, Issue 10, pp. 1568–1579, 2005.

[11] A. Schmidt, F. Lafarge, C. Brenner, F. Rottensteiner, and C. Heipke, "Forest point processes for the automatic extraction of networks in raster data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 126, pp. 38–55, 2017.

[12] T. Li, M. Comer, and J. Zerubia, "A Connected-Tube MPP Model for Object Detection with Application to Materials and Remotely-Sensed Images," in *IEEE International Conference on Image Processing (ICIP)*. Oct. 2018, pp. 1323–1327, IEEE.

[13] V. Mnih, *Machine Learning for Aerial Image Labeling*, Ph.D. thesis, University of Toronto, 2013.

[14] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," *IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, 2017.

[15] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *ICLR*, 2016.

[16] A. Mosinska, P. Márquez-Neila, M. Kozinski, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3136–3145.

[17] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, Issue 3, pp. 347–359, 2009.

[18] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the hsv color space for image retrieval," in *Proceedings. International Conference on Image Processing*, Sep. 2002, vol. 2, pp. II–II.

[19] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, and A. Vanhoutte, "Similarity measures in scientometric research: The jaccard index versus salton's cosine formula," *Inf. Process. Manage.*, vol. 25, no. 3, pp. 315–318, May 1989.