



Urban Streetscape Tree Density Estimation Algorithm Based on Image Semantic Segmentation

Bin Wang¹, Ping Sun¹, Zhongwang Zhang², and Lin Ma¹ (✉)

¹ School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

malin@hit.edu.cn

² China National Aeronautical Radio Electronics Research Institute, Shanghai, China

Abstract. In the aspect of urban forest density estimation, there is a lack of automatic or efficient estimation methods. For the existing research on urban streetscape trees, mainly uses lidar to process point cloud data or combines deep learning to achieve tree segmentation and detection. However, these methods lead to too much computation, low efficiency, and fail to provide estimation results of urban tree density. By processing the image data, this paper proposes a tree density estimation algorithm based on image semantic segmentation, which deals with the only image data in the whole process, and realizes the estimation of tree density in the city streetscape. This algorithm is more efficient and accurate than the complex point cloud operation or the method combining point cloud with a deep learning algorithm.

Keywords: Semantic segmentation · Urban streetscape · Tree density estimation

1 Introduction

Trees growing in cities or large areas of green forest can be uniformly defined as urban forests. The existing database of urban forests, however, there are insufficient data, updating the problem of not in time, the traditional manual measurement and estimation cannot meet the needs of modern urban forest estimation, people are eager to realize automation and efficient estimation in urban forest density estimation, in order to facilitate real-time monitoring of their time and space distribution.

The diverse composition of cities makes it challenging to use lidar filtering. At the same time, complex 3D point cloud computing makes it more difficult to achieve segmentation and 3D reconstruction of urban trees. In [1], voxel grid filtering is used to reduce the number of point clouds. In this algorithm, grid filtering, spatial clustering, and the k-nearest neighbor method are used in turn to extract the final tree height. The effect is good, but the algorithm design is still complex. In [2], A new method based on point density is proposed to reconstruct the complete three-dimensional structure of a single tree, but this algorithm is very dependent on the composition of the environment,

and if it is too complex, the classification accuracy and reconstruction accuracy will be reduced. With the development of deep learning, tree detection and tree segmentation are realized by combining LIDAR with the target detection algorithm in [3]. The overall segmentation accuracy is very high, but there is still a lack of in-depth research on the statistics of trees. In [4], the LiDAR point cloud was used to detect a single tree and its related measurements, and the measurements of tree height, canopy diameter, and depth were realized. However, there was no statistical study on the distribution density of urban forests. In terms of computational complexity, the use of lidar data combined with deep learning algorithms will be more complex.

Aiming at these problems, this paper proposes an estimation algorithm of urban tree density based on image semantic segmentation. Firstly, we segment the whole image. Secondly, the trees' areas are extracted through a series of image processing techniques, which reduces the number of image pixels that the algorithm needs to traversal in the process of estimating the density of trees. Finally, the size of the tree pixel in the whole image is calculated by the number of pixels to realize the estimation of the density of trees. The algorithm does not use three-dimensional data similar to the lidar point cloud but achieves estimated extraction of tree density in the scene through segmentation and processing of the whole image. Therefore, compared with the processing algorithm of point cloud data, its complexity is reduced and it only needs to realize image processing.

The chapters of this paper are arranged as follows. In Sect. 2, the basic principle of image semantic segmentation is briefly introduced. Section 3 introduces an algorithm of urban tree density estimation based on image semantic segmentation. Section 4 gives the overall process of algorithm implementation and the final detection results. The experimental results are analyzed and the advantages of the proposed algorithm are given. Finally, the thesis is summarized in Sect. 5.

2 System Design

Since the appearance of the FCN network [5] in 2014, image semantic segmentation has realized pixel-level image segmentation. It is an end-to-end neural network based on CNN structure. In the input layer, it is not restricted to the size of the image and the final full connection layer of CNN is removed, making the semantic segmentation using a pre-trained network more efficient. A jump connection is introduced to solve the problem of pixel location with rough up-sampling. But it is insensitive to the details of the image, which restricts the application of data requiring precise statistics such as tree density. After FCN came into being, various forms of neural networks came into being [6–9]. In 2020, the University of Science and Technology of China and Microsoft jointly proposed a High-Resolution Net (HRNET) [10]. The network maintains a high resolution in the whole process. Starting from the high-resolution convolution stream, it gradually increases the high-resolution convolution stream to the low-resolution convolution stream and connects the multi-resolution convolution stream in parallel. The advantage of HRNET is that in terms of feature extraction, image details can always be kept without loss, which is richer in semantic expression and more accurate in space.

Therefore, this paper proposes an urban streetscape tree density estimation algorithm based on HRNET semantic segmentation. Firstly, the open-source cityscapes images are

selected as the original input of the whole algorithm, and the semantic region of trees in the image is extracted by using the HRNET network. The extracted results are processed by image grayscale, edge extraction, mathematical morphology, to achieve the extraction of tree contour. Finally, count the number of tree pixels in the region and get the final estimation of urban tree density according to the number of pixels in the whole image. The system block diagram of the whole algorithm is shown in Fig. 1.

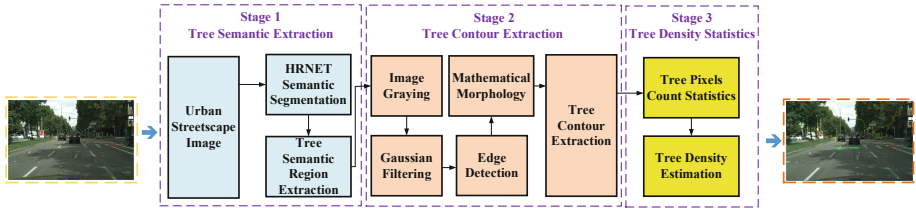


Fig. 1. Block diagram of tree density estimation system based on image semantic segmentation

In Fig. 1, a set of open-source cityscapes image data sets are used as the input of the entire tree semantic extraction and tree density estimation system, and the final output results are used to achieve tree density estimation for each image. Among them, the estimated algorithm of urban tree density based on HRNET semantic segmentation reduces the computational complexity of the original laser radar combined with image processing, improves the computing efficiency of urban tree density estimation, and realizes the automatic tree density estimation. In the algorithm proposed in this paper, through a series of image processing methods, HRNET semantic segmentation results can not only focus on the tree region but also make full use of the pixel number of the region to achieve automatic urban tree density estimation. So that the estimated algorithm used in the actual application has a better real-time interaction.

3 Tree Density Estimation Algorithm

3.1 Semantic Region Extraction Based on HRNET

HRNET network is used for semantic segmentation to achieve segmentation of each type of object in the city streetscape, such as road surface, trees, etc. Through segmentation, each type of object can be analyzed in detail, which lays a foundation for estimating of the density of trees in the city streetscape. Its structure is shown in Fig. 2.

It can be observed that the network is divided into four stages, each of which is composed of channel maps with different resolutions, and each of which has one more branch than the previous one. The new branch is the result of the step convolution fusion of all the feature images of the previous stage. The resolution at the end of the stage will be half of the resolution of the previous branch, and the number of channels will be doubled. On the left is the input image which can be extracted from the semantic region of the city streetscape through the HRNET network.

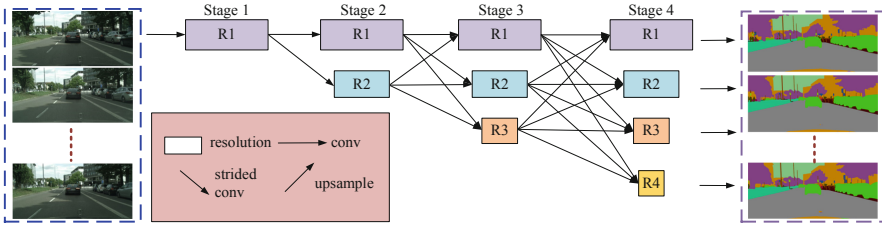


Fig. 2. HRNET network structure

3.2 Tree Contour Extraction Based on Semantic Region and Image Processing

The second step of the algorithm is based on semantic region and image processing to complete the tree contour extraction, mainly through image grayscale processing, Gaussian filtering, edge detection, and mathematical morphology processing.

(1) Image grayscale

The semantic segmentation map of urban trees is grayed to keep only the urban tree area that we are interested in and remove the redundant other areas. The whole image is traversed, and the tree area is assigned a value of 255, that is, white, while the other areas are assigned a value of 0. It can also be processed by the grayscale formula as:

$$\text{Gray} = 0.299R + 0.587G + 0.114B \tag{1}$$

where, RGB represents the image's three channels, red, green, and blue.

(2) Gaussian filtering

The Gaussian filter is a linear filter, which can eliminate the noise generated in the process of digital processing and extract useful features in the image. For the next edge detection, an appropriate Gaussian filter should be selected to suppress a kind of noise whose probability density function follows Gaussian distribution, and the Gaussian kernel size of 7×7 is selected. The calculation formula of Gaussian kernel parameters by second-order Gaussian function is shown as follows:

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{2}$$

where, x^2 and y^2 respectively represent the distance between the pixels in the Gaussian core and the center pixel, and represent the standard deviation. As the standard deviation increases, the effect of Gaussian filtering will be more obvious.

(3) Edge detection

The canny operator is used to realize edge detection. It adopts the method of smoothing first and then derivative. It belongs to the first-order differential filter and is an optimization operator that can achieve a strong filtering effect. By calculating the gradient of the image, the amplitude and angle of the edge of the image are obtained and the change of the gray value is detected. If a drastic change is found somewhere, the area can be determined as the edge of an object. The Sobel operator

is selected to obtain the pixel gray value corresponding to the 8-neighborhood region. The template operator and the pixel gray value of 8-neighborhood are shown in (3) and (4) respectively.

$$M_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad M_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

$$g_{px} = \begin{bmatrix} f(x-1, y+1) & f(x, y+1) & f(x+1, y+1) \\ f(x-1, y) & f(x, y) & f(x+1, y) \\ f(x-1, y-1) & f(x, y-1) & f(x+1, y-1) \end{bmatrix} \quad (4)$$

where, M_x is the x direction template, M_y is the y direction template.

Horizontal and vertical templates are shown in formula (5), and pixel gray value and angle in 8-neighborhood are shown in formula (6) and (7) respectively.

$$\begin{aligned} G_x &= f(x+1, y+1) - f(x-1, y+1) + 2f(x+1, y) \\ &\quad - 2f(x-1, y) + f(x+1, y-1) - f(x-1, y-1) \end{aligned} \quad (5)$$

$$\begin{aligned} G_y &= f(x-1, y+1) - f(x-1, y-1) + 2f(x, y+1) \\ &\quad - 2f(x, y-1) + f(x+1, y+1) - f(x+1, y-1) \end{aligned}$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (6)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (7)$$

(4) Mathematical morphology processing

Mathematical morphology processing is used to further extract the boundary of the tree area, where corrosion and expansion are two basic operations. $B_{(x,y)}$ is the convolution template, B_x is the structural elements, and A is the input image, then the corrosion of B_x to A is shown as follows:

$$A - B_x = \{f(x, y) | B_{(x,y)} \subseteq A\} \quad (8)$$

where, $f(x, y)$ is the pixel value of any pixel point in the image A . Similarly, the expansion of B_x with respect to A is shown as follows:

$$A + B_x = \{f(x, y) | B_{(x,y)} \cap A = \emptyset\} \quad (9)$$

where, \emptyset is the empty set.

3.3 Tree Density Estimation Algorithm Based on Tree Contour Region

The density of an object in an image is defined as the proportion of the number of pixel points of the object to the total number of pixel points in the whole image. Therefore, the density calculation of trees can be expressed as follows:

$$\rho_{\text{tree}} = \frac{n_{\text{tree}}}{M \times N} \quad (10)$$

where, ρ_{tree} is the density of trees; n_{tree} is the number of pixel points in the tree part; and M, N are the height and width of the whole image respectively. To calculate the number of pixels in the tree part, we only need to traverse the whole image and then sum the number of pixels in the tree part.

4 Experiment and Analysis

4.1 Experimental Environment

The implementation of this algorithm is based on Linux-Ubuntu operating system and PyTorch framework under a deep learning environment. The open-source data sets Cityscapes are selected. The data set LeftImg8Bit and GtFine store image sets and label sets respectively, and each folder contains training sets, verification sets, and test sets respectively. A total of 5,000, 2,975 for training, 500 for verification, and 1,525 for testing. To reduce the time spent on training the network, 1100 pictures of Stuttgart city were selected for this project, and 8 pictures were selected as shown in Fig. 3. The semantic segmentation graphs were selected as in Fig. 4.



Fig. 3. Dataset of Stuttgart

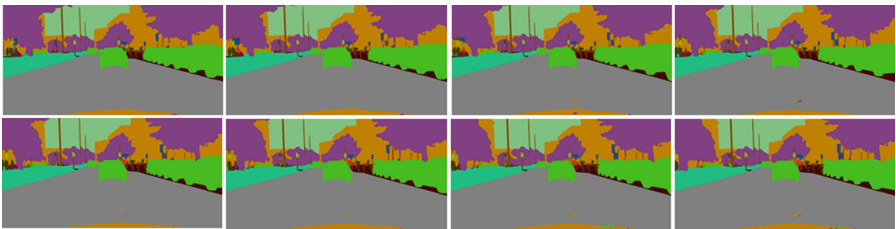


Fig. 4. Semantic segmentation of Stuttgart

4.2 The Experimental Results

According to the results of urban tree distribution in Stuttgart, the criteria for estimating the density of trees in the city can be set, as shown in Table 1. Generally, the density of trees in the city can be divided into three levels: low, medium, and high.

Table 1. Standard for estimating urban tree density

The serial number	Tree density range	The thick degree
1	[0, 0.18)	Low
2	[0.18, 0.28)	Medium
3	[0.28, 1]	High

A total of 6 pictures of 270–280 frames are selected with the span of every two frames to estimate the density of trees. The peripheral contour of trees is marked in yellow and brown. The value of tree density and the degree of trees in each frame of the picture is output in the middle of the picture. The six pictures are shown in Fig. 5.



Fig. 5. Tree extraction and tree density output

Six pictures are selected in the figure above. It can be seen from the analysis that as the vehicle progresses, it generally moves towards the direction of dense distribution of trees, so the value of tree density continues to increase. Meanwhile, the overall value is greater than 0.28, so the thickness of the trees is shown as high (Table 2).

Table 2. Tree density and dense degree table of 270–280 frames

The serial number	Frame	Tree density	The thick degree
1	270	0.289906	High
2	272	0.292167	High
3	274	0.292631	High
4	276	0.294183	High
5	278	0.295065	High
6	280	0.296948	High

5 Conclusion

Aiming at the problems of low efficiency of the algorithm of processing point cloud data by lidar and the lack of automation of urban tree density estimation, this paper proposes an urban streetscape tree density estimation algorithm based on image semantic segmentation. Firstly, the open-source cityscapes data set combined with HRNET semantic segmentation network is used to divide the semantic regions of trees. Secondly, a series of image processing algorithms are used to extract the contours of trees. Finally, the density of trees in the city is estimated based on the statistics of pixels. The whole process is based on image data processing, without the need for other sensor collection point cloud processing or combined with a deep learning algorithm to achieve tree region segmentation, the algorithm enhances the real-time interactivity of tree density estimation, so that the density of trees with faster speed and automation.

Acknowledgment. This paper is supported by National Nature Science Foundation of China (41861134010, 61971162) and National Aeronautical Foundation of China (2020Z066015002).

References

1. Liu, Y., Xing, M., Zhou, X., et al.: Tree height extraction in sparse scenes based on UAV remote sensing. In: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 6499–6502 (2020)
2. Xiangyu, W., Donghui, X., Guangjian, Y., et al.: 3D reconstruction of a single tree from terrestrial LiDAR data. In: 2014 IEEE Geoscience and Remote Sensing Symposium, pp. 796–799 (2014)
3. Alon, A.S., Festijo, E.D., Juanico, D.E.O.: Tree detection using genus-specific retinanet from orthophoto for segmentation access of airborne LiDAR data. In: 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), pp. 1–6 (2019)
4. Tarawally, M., Wenbo, X., Weiming, H., et al.: Effect of deforestation on land surface temperature: a case of Freetown and Bo town in Sierra Leone. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 5232–5235 (2018)
5. Long, J., Shelhamer, E., Darrell, T., et al.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
6. Badrinarayanan, V., Kendall, A., Cipolla, R., et al.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
7. Chen, L., Papandreou, G., Kokkinos, I., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018)
8. Yu, C., Wang, J., Peng, C., et al.: Learning a discriminative feature network for semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1857–1866 (2018)
9. Yu, C., Wang, J., Peng, C., et al.: BiSeNet: bilateral segmentation network for real-time semantic segmentation. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
10. Wang, J.D., Sun, K., Cheng, T.H., et al.: Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3349–3364 (2020)