IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9324487

DEEP LEARNING FOR VEGETATION IMAGE SEGMENTATION IN LAI MEASUREMENT

Cunshi Ma[1], Yunping Chen[1,], Lei Hou[1], Baihui Li[1], Yan Chen[1], Yuan Sun[2,*], Xingfa Gu[2]*

[1]School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China.
[2] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, 100094, China.
[*]E-mail: chenyp@uestc.edu.cn; sunyuan@radi.ac.cn

ABSTRACT

For the measurement of LAI (Leaf Area Index) by DHP (Digital Hemispherical Photography) method, imprecise segmentation is the key error source. In this paper, to our knowledge, a deep learning algorithm is used for the first time to segment upward hemispherical image of vegetation. Pix2pix, a general mapping learning model, was improved in our study to make it more suitable for processing segmentation problem. Thousands of images collected in the field were labeled to train the model, and the conventional methods based on pattern recognition, such as the Otsu and HSV, were compared. The result shows that the improved pix2pix algorithm significantly improved the accuracy of the segmentation, which reached to 0.9834. Furthermore, this model has a good performance in processing pictures of complex environments, and the segmentation of edge details has also been optimized. Those results show that the method has great potential to improve the LAI measurement accuracy.

Index Terms— *Leaf area index, image segmentation, pix2pix*

1. INTRODUCTUION

The leaf area index is the half of the total leaf area per unit area of vegetation[1] and is an important parameter in agricultural science, environmental science, and remote sensing science[2]. For passive optical instrument, there are generally two methods for LAI measurement. One is to measure the radiation transmittance through sensors, which is adopted by the LAI2200 series of instruments, and the other is to use an optical camera to take pictures of vegetation and then perform image analysis, which is adopted by the CI-110 instrument. Relatively, the latter is low cost, easy to upgrade and convenient for networking, thus, it is widely used for ground measurement and remote sensing validation[3]. Meanwhile, a critical step of this method is to separate the leaves from the background.

The result of image segmentation is an important factor that affects the accuracy of the final LAI result. Although conventional segmentation methods (such as Otsu method and HSV threshold segmentation method) have good results in some specific weather conditions, limitations still exist in processing vegetation image segmentation. Under the influence of strong light or other weather, the captured image is overexposed, or the background is covered with white clouds. For these harsh situations, conventional segmentation methods can not get distinct segmentation results. Therefore, the deep learning method is applied to segment vegetation images and the improved pix2pix model is used for training and prediction.

2.MEASUREMENT AND SEGMENTATION THEORY

2.1. DHP measurement theory

Digital hemispherical photography method infers LAI from measurements of light transmission through canopies with the use of fisheye camera. Firstly, the upward hemispherical picture is taken under the vegetation, and then the picture is divided into two parts: vegetation and background. An important parameter, the ratio of the leaves pixel number to the total pixel number at a specific field of view angle, is used to calculate the LAI by formula, such as the widely used Beer-Lambert law[3, 4]. A large number of experiments have proven that under the premise of distinct image segmentation, for low vegetation, the LAI value measured based on Beer-Lambert law is extremely accurate. Therefore, the accuracy of DHP method is mainly influenced by the accuracy of image segmentation.

2.2. Pix2pix model theory

Image segmentation could also be regarded as a type of image mapping problem. Among the deep learning models, Pix2pix is a general solution to image mapping problems. It can get a good prediction model without proposing a specific loss function for a specific problem. Pix2pix is inspirited by the ideas of cGAN, but is different from it. When cGAN is inputting data into the generate network, it will not only input noise, but also a condition. The generated image will be affected by specific conditions. Then if an image is used as a condition, the generated fake image has a corresponding

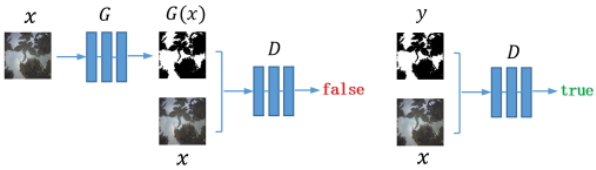

Fig.1. Training procedure of pix2pix

relationship with the input image, thereby implementing an image mapping process. The training procedure of pix2pix is shown as figure 1.

For the generator, the "U-Net" architecture is used. The input image *x* is encoded and then decoded into image *G(x)*; for the discriminator, the convolutional "PatchGAN" classifier is used. Under the condition of input image *x*, the generated image *G(x)* is judged to be false, and the real picture *y* is judged to be true, and the model only penalizes the image on a small size[5].

2.2.1. Loss function

The objective function of cGAN is as follows:

$$L_{cGAN}(G,D) = E_{x,y}[logD(X,Y)] + E_{x,z}[\log(1 - D(x,G(x,z)))]$$

Where G tries to minimize the goal and D tries to maximize the goal, that is:

$$G^* = \arg minG \ \ maxD \ \ L_{cGAN}(G,D)$$

For image translation tasks, G's input and output actually share a lot of information. Therefore, in order to ensure the similarity between the input image and the output image, L1 Loss is also added:

$$L_{L1}(G) = E_{x,y,z}[||y - G(x,z)||_1]$$

That is, the L1 distance between the generated fake images and the real images ensures the similarity of the input and output images.

So the final loss function is:

$$G^* = \arg minG \ \ maxD \ \ L_{cGAN}(G,D) + \lambda L_{L1}(G)$$

2.2.2. Improved generate network structure

For the image-to-image translation task, the input and output are different, but the two should share some information. Therefore, the structure in the input is roughly aligned with the structure in the output. The structure of the generated network is shown in Figure 2.

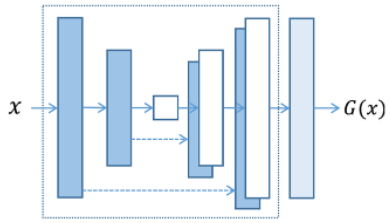

Fig.2. Improved generate network structure

The U-Net structure is based on the Encoder-Decoder model, while encoder and decoder are symmetrical structures. The difference between them is U-Net connects the i-th layer and the n-i layer, where n is the total number of layers. This connection method is called skip connection. The image sizes of the i-th layer and the n-i layer are the same, so that they can carry similar information, thereby retaining more information about the image structure and details. In addition, the pix2pix model is a pixel learning mapping model. The predicted results are not binary images, it will have pixel values near 0 and 255, so the transition edges of the predicted image will look coarse and blurred. In order to solve this problem, the mean-field approximation algorithm is used, the dense CRF is decomposed into simple calculations, and each step is described as a convolution operation[6]. Then add the convolutional layer behind the U-Net network structure to form a complete end-to-end structure. It will reduce noise, make the output result have a clearer outline, and improve segmentation quality.

2.2.3. Discrimination network structure

If the typical L1 and L2 are used as the loss function of a single mapping problem, the reconstructed image is very blurred, and it cannot recover the high frequency part of the image well. In order to judge the local part of the image better, the Pix2pix discrimination network uses the "PatchGAN" structure, which divides images into multiple fixed-size patches, and the trueness and falseness of each patch are judged independently. Finally, the average value is used as the output value.

3. EXPERIMENT

Thousands of fisheye pictures were taken below vegetation in Guangzhou, Nanjing, Chengdu, Changchun and other places. These pictures include rice, corn, soybean, Bush, grass, tea tree, zucchini and other plants. A 1024×1024 picture was cut at the center of each fisheye picture, and then was compressed into a 512×512 picture. The human-computer interaction labeling software written by ourselves was used to get the segmentation result, that is, the training label. Because the number of valid samples is limited, each picture and its label were divided into four 256×256 pictures. Then we perform image enhancement operations such as rotating the angle and adjusting the brightness of the picture. A total of more than two thousand training samples were obtained. 80% of the samples were randomly selected for model training, 10% for validation and 10% for testing.

For the test samples, the most widely used Otsu threshold method and HSV threshold segmentation method were used for segmentation, and they were compared with improved pix2pix model inference method. Because the result image has only two data, 0 and 255, there is no need to design a complex evaluation method. The label of the test picture is the standard, and the average segmentation accuracy and SSIM of each segmentation method is calculated, and then the precision and recall are used to evaluate the model performance.

4. RESULTS AND ANALYSIS

Several pictures of different vegetation were selected for display. The results of three segmentation methods are shown in Figure 3, Figure 5-7.

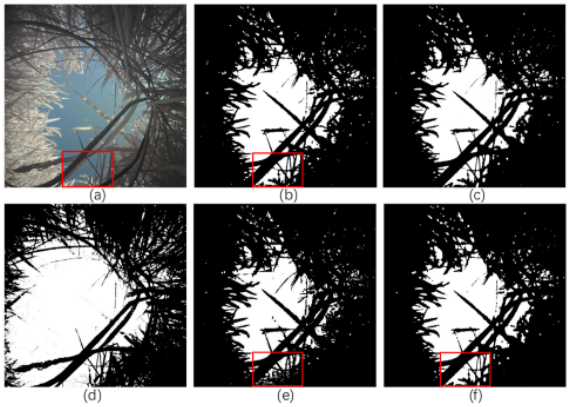

Fig.3. Rice segmentation results: (a)the original image; (b)the label; (c)the pix2pix model prediction result (d)the Otsu segmentation result; (e)the HSV threshold segmentation result; (f)the improved pix2pix model segmentation result(Figure 5-7 has the same structure, and is not annotated)

Figure 3(a) is a picture of rice. An important feature of this type of picture is that some rice leaves are brighter than the sky, so the pixels in the picture will be divided into three categories : dark leaves, sky, bright leaves. Figure 3(d) is the result of adaptive threshold Otsu segmentation. Bright leaves and sky are classified into one category, and dark leaves are classified into one category.

Fig.4. Background color in different environments

For the same shooting position, the pixel value will change greatly under the influence of weathers. As shown in Figure 4, the color of sky is inconsistent. The accuracy of the HSV threshold segmentation method depends on whether the empirical value is applicable to the picture. After counting a large number of pictures, a threshold is determined which is effective for most pictures. Figure 3(e) is the result of HSV threshold segmentation, which can basically separate the background from the leaves, but misrecognize the dark sky. Figure 3(f) shows the reasoning result of the improved pix2pix model. It is obviously better than the Otsu method and the HSV threshold segmentation method.

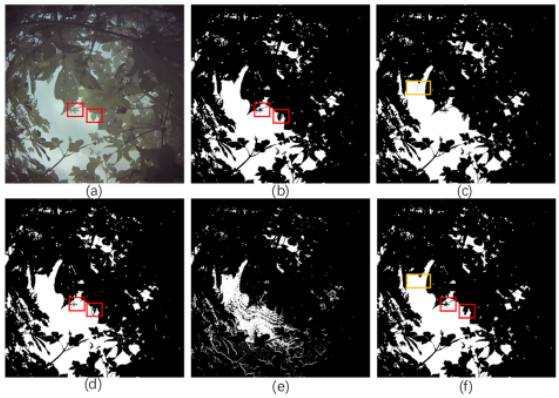

Fig.5. Bush segmentation results

Figure 5(a) is an original picture of low bushes. The Otsu segmentation method is very suitable for this type of picture with the weather of cloudy. As we can see from figure 5(d), the result looks distinct. But the HSV threshold segmentation result is fuzzy. Comparing the result of the improved pix2pix model and the Otsu segmentation method, the former handles the details better than the latter. During the training process, we found a particularly interesting and worthy research question. At first, for this kind of picture, we used the Otsu segmentation result as the training label. The result predicted by the training model is shown in Figure 5(b), but the detail is better than the original label. Then we use the result of the initial prediction as the label to retrain the model. Generally speaking, the model obtained in this way has better performance than the model trained for the first time.

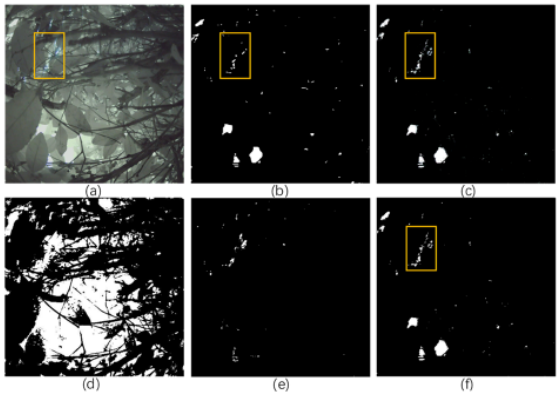

Fig.6. Tea tree segmentation results

Figure 6(a) is a picture of a tea tree. The camera is under dense vegetation, almost all the field of vision is filled with leaves. The leaves are divided into two parts by the Otsu method, which is obviously not desirable. The Pix2pix model can roughly reflect the characteristics of the picture, but purple and blue dots appear after the image is enlarged. The improved pix2pix model solves this problem which is caused by the innate generated structure.

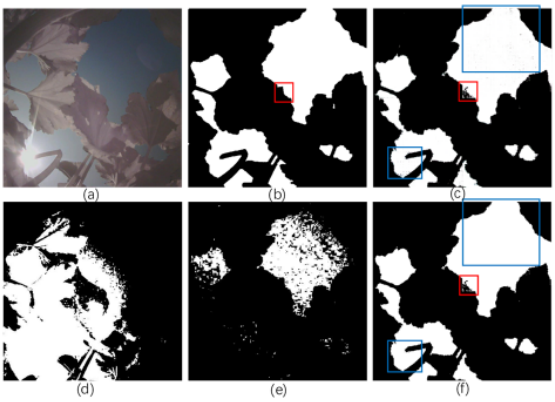

Fig.7. Zucchini segmentation results

Figure 7(a) is the zucchini leaves. It can be seen that, under the influence of strong sunlight, the Otsu method and the HSV segmentation method do not work, and the trained pix2pix model is basically unaffected. This reflects that pix2pix model segmentation is still effective in complex background environments. In the past, in order to ensure the quality of segmented pictures and the accuracy of LAI value, the LAI measurement experiment basically chooses to collect pictures on cloudy days or in the morning and evening. The application of deep learning methods to digital hemispherical photogrammetry will break time and environmental constraints.

Figure 7(c) and 7(f) are stitched from four predicted pictures. Because the model has insufficient processing capabilities for the corners, some of the pictures have poor segmentation of the corner regions like in the red box. In the blue box area, we can see that there are a lot of noises in the prediction result of the pix2pix model. The pixel values of the three channels of these noises are close to 255. The improved pix2pix model has an end-to-end structure, which eliminates the noise and completes the process of mapping to segmentation.

Through statistics on all test samples, we get the average segmentation accuracy and SSIM of the images as shown in Table 1:

Tab.1. Average accuracy and SSIM of three segmentation methods

Segmentation method	OTSU	HSV threshold	Improved pix2pix
Average accuracy	0.8927	0.8142	0.9834
Average SSIM	0.7724	0.6342	0.9030

In addition, we calculated the precision and recall of leaves and background respectively, and the data are shown in Table 2:

Tab.2. Performance of the pix2pix model on the test set

	Precision	Recall
Leaves	0.9874	0.9846
Background	0.9174	0.9283

As shown in the tables, the accuracy and SSIM of improved pix2pix inference method is far higher than two conventional segmentation methods. Analyzing the data set, the area of the leaves in most pictures is higher than the sky, so the precision and recall of the background are smaller than those of the leaves. Just considering the accuracy and recall of the background, the performance of this model on the data set is also fine. A single conventional method cannot segment all kinds of vegetation pictures, and multiple conventional methods cannot be used together effectively. It is difficult to know which method is suitable for a picture. However, the improved pix2pix model has a strong applicability.

5. CONCLUSIONS

For the LAI measurement solution, we applied the deep learning method to the digital hemispherical photogrammetry for the first time, and used the improved pix2pix model for image segmentation, which greatly improved the accuracy of segmentation and further improved the accuracy of LAI measurement. However, the model's significant segmentation accuracy has a dependence on the samples, enough samples of various vegetation types should be collected and processed to make it have higher application value.

6. ACKNOWLEDGMENTS

This work is supported by the Sichuan Science and Technology Plan Project (No. 2019YJ0201) and the Common Application Support Platform for Land Observation Satellite of National Civil Space Infrastructure （No. Y930280A2F).

7. REFERENCES

[1] K. S. Fassnacht, S. T. Gower, J. M. Norman, and R. E. Mcmurtric, "A comparison of optical and direct methods for estimating foliage surface area index in forests," *Agricultural & Forest Meteorology,* vol. 71, no. 1-2, pp. 183-207, 1994.

[2] K. Yan *et al.*, "Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison," *Remote Sensing,* vol. 8, no. 6, Jun.2016.

[3] G. J. Yan *et al.*, "Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives," *Agric. For. Meteorol.,* vol. 265, pp. 390-411, Feb.2019.

[4] T. Nilson, " A theoretical analysis of the frequency of gaps in plant stands," *Agric. Meteorol,* vol. 8, pp. 25-38, 1971.

[5] J. Z. P. Isola, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017,* pp. 5967-5976, 2017.

[6] S. Zheng *et al.*, "Conditional random fields as recurrent neural networks," *15th IEEE International Conference on Computer Vision, ICCV 2015,* pp. 1529-1537, 2015.