A Method of the Coverage Ratio of Street Trees Based on Deep Learning

Wen Han, Lei Cao, Sheng Xu*

Nanjing Forestry University, Nanjing (China)

Received 15 November 2021 | Accepted 3 June 2022 | Published 25 July 2022

Abstract

The street trees coverage ratio provides reliable data support for urban ecological environment assessment, which plays an important part in the ecological environment index calculation. Aiming at the statistical estimation of urban street trees coverage ratio, an integrated model based on YOLOv4 and Unet network for detecting and extracting street trees from remote sensing images is proposed, and obtain the estimated street trees coverage ratio in images accurately. The experiments are carried out under self-made dataset, and the results show that the accuracy of street trees detection is 94.91%, and the street trees coverage ratio is 16.30% and 13.81% in the two experimental urban scenes. The MIoU of contour extraction is 98.25%, and the estimated coverage accuracy is improved by 6.89% and 5.79%, respectively. The result indicates that the proposed model achieves the automation of contour extraction of street trees and more accurate estimation of street trees coverage ratio.

Keywords

Ecological Environment Index, Object Detection, Remote Sensing Image, Street Trees Coverage Ratio, Unet, YOLOv4.

DOI: 10.9781/ijimai.2022.07.003

I. Introduction

THE coverage ratio of street trees is an important indicator affecting the urban ecological environment index [1], which can provide reliable data support for urban ecological environment assessment. With the development of deep learning [2], the method of estimating the coverage ratio of street trees in remote sensing images based on deep learning can reduce manual intervention, improve the efficiency of measurement and provide more accurate data information to relevant administrative departments.

In terms of image processing [3], the convolutional neural network [4] (CNN) is widely used. Among the field of image processing methods, CNN is the most outstanding model in deep learning. It is a special multi-layer perceptron designed to detect two-dimensional images. The essence of the convolutional kernel is a feature extractor. CNN has the structural characteristics of local perception region, weight sharing and pooling. Local perception region and weight sharing significantly reduce the number of parameters of CNN and improve the network performance. The spatial subsampling enables CNN to have a certain scaling and translation invariance and has a stronger generalization ability. CNN incorporates feature extraction into model learning, organically combines feature learning with classification learning, and realizes image interpretation more effectively.

Today, there are few studies on the estimation of street trees coverage ratio in remote sensing images based on deep learning in the domestic and overseas. However, there are still some scholars at home and abroad who have studied the street greening ratio based on different streetscape data platforms. For example, Long Ying [5] has studied street greening in 245 major cities in China based on Tencent Streetscape Platform; Li Xiaojiang et al [6]. modified the green landscape index by combining it with Google Streets View, and achieved the automatic measurement of urban street green ratio; Seiferling et al. [7] quantified street trees in New York and Boston based on Google Street View images. Yan Li et al. [8] proposed a vegetation extraction method based on attention model in remote sensing image, under different environmental topographic and climatic conditions, vegetation differentiation and symbiosis with other land features can separate a single vegetation. Although the above researches reflect the coverage ratio of street trees to a certain extent, the street scenery taken from different angles will have geometric deformation, and the observation information of images is limited and hard to be used for accurate measurement and calculation. However, observation information of remote sensing images is more macroscopic and accurate, and we can get more detailed messages from the given images, so, it is beneficial for us to use remote sensing images to study street trees coverage ratio. There are also many challenges, such as shading between trees, insufficient lighting, and weather and seasonal factors that cause errors in estimates of street trees coverage ratio.

Contributions of our work are as follows:(1) YOLOv4 [9] is used to detect the street trees in remote sensing images, and the coverage ratio of the street trees is estimated through the coordinate information of the bounding box; (2) an integrated model of remote sensing images object detection and contour extraction is proposed, which is an end-to-end process [10]. While calculating the street trees coverage ratio through this model, the problems arising in estimation based on object detection model and instance segmentation can be effectively solved, and the estimated value is closer to the real value.

The structure of this document is as follows: section II gives introduction about related algorithms to this work, section III provides experimental preparation, including dataset, hardware environment and evaluation index, section IV presents the experiment results, gives full discussion about the varieties of cases what occur in the

* Corresponding author.
E-mail address: xusheng@njfu.edu.cn

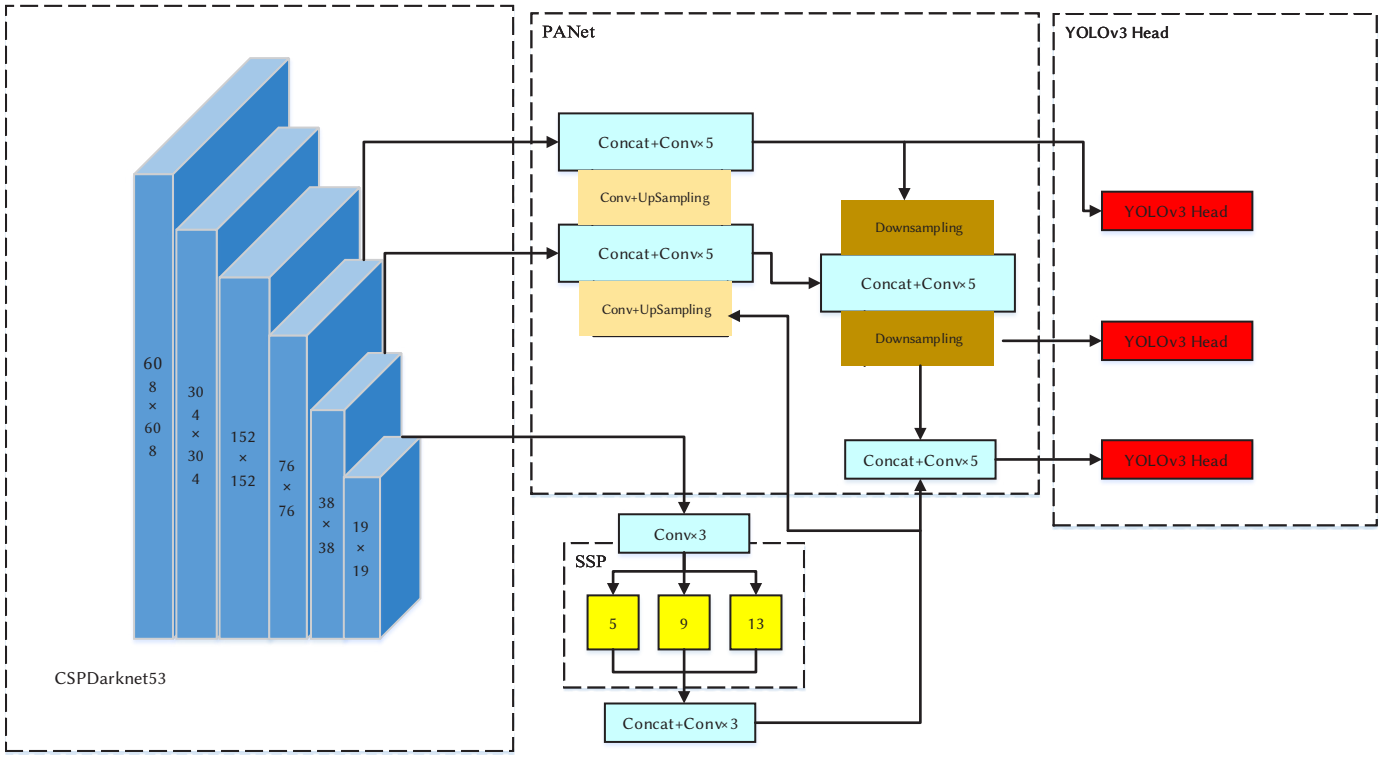

Fig. 1. The network structure of YOLOv4.

experiment, and carry out a comparative experiment among some mainstream algorithms to prove the superiority of the proposed algorithm, Section V makes a summary of the work and puts forward the vision of future work.

II. Related Algorithms

A. The Structure of YOLOv4

Object detection is an important application of artificial intelligence, which is to identify the object in the image and mark the position of the object. Here we used YOLOv4 to detect the street trees in the remote sensing images.

By introducing Mosaic and GA [11] to select the optimal hyper-parameters, YOLOv4 improves the existing methods to a lower training threshold, which can achieve better results under the limited GPU resources. The network structure of YOLOv4 is shown in Fig.1 below. CSPDarknet53 is the backbone, SPP [12] (Spatial Pyramid Pooling) is the additional module of neck, PANet [13] (Path Aggregation Network) is the feature fusion module of neck, and YOLOv3-head [14] is the head. CSPDarknet53 added CSPNet (Cross Stage Partial Network) to each large residual block of Darknet53 and integrated it into the feature map through gradient changes. The feature map was divided into two parts, one of which was convolutional operation. The other part is combined with the last convolution. PANet makes full use of feature Aggregation. The fusion method is changed from addition to multiplication, which makes object detection capability more accurate. In order to get the extraction of the contour of the street trees, we introduce Unet [15] to achieve this goal.

B. Unet Network Model

Image segmentation algorithms based on deep learning mainly have two core frameworks: one is image feature extraction based on CNN, and the other is the upsampling/deconvolution segmentation framework based on global neural network, such as FCN [16] (Fully Convolutional Networks). The former cannot be accurately segmented because the category probability of each pixel cannot be identified by the full connection layer. The latter changes the full connection layer of the former to the convolution layer, and introduces the upsampling before multiple pooling operations, which solves the problem of accurate segmentation, but the effect of edge extraction is not good enough.

Unet draws on the characteristics of FCN. The network structure consists of two symmetric parts: contracting path and expanding path. Contracting paths adopt 3×3 convolution and pooled down-sampling, which can obtain shallow features and deep features, then capture the relationship among pixels. The 3×3 convolution and upsampling are used in expansion paths. While upsampling, the deep features and shallow features are combined in a cascading way to obtain the accurate position of the image to be segmented. The network structure of Unet is shown in Fig.2 below.

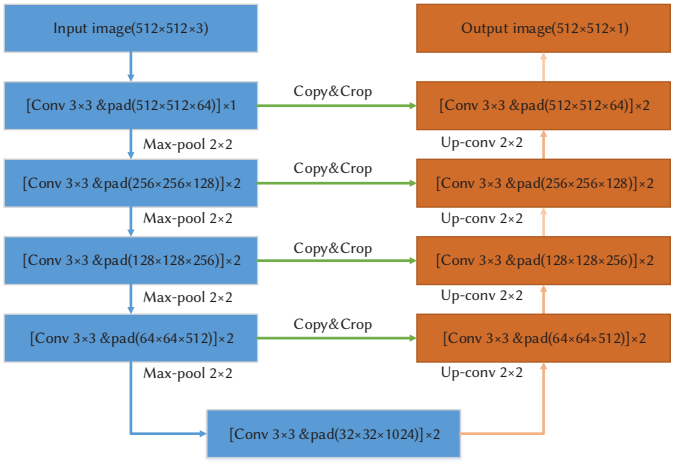

Fig. 2. The network structure of Unet.

C. Integration of Object Detection and Contour Extraction From Images

Based on deep learning, an integrated process of object detection and contour extraction in remote sensing image is proposed. This is an end-to-end process. Firstly, according to the detection results of YOLOv4, the coordinate information and bounding box are acquired, and the segmentation image of the corresponding region is obtained by Unet. Finally, the obtained segmentation image of the corresponding region is fused with the corresponding region of the object detection to get the final result of object detection and contour extraction. As is shown in Fig.3.

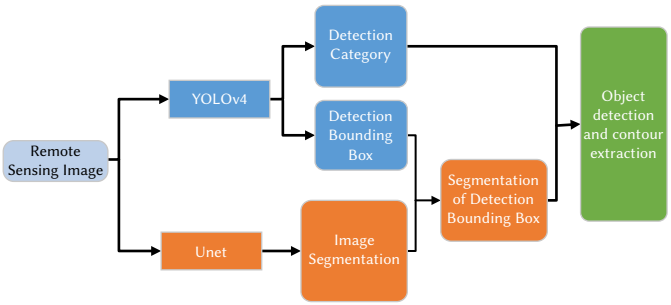

Fig. 3. Integrated flow chart of object detection and contour extraction.

III. Experiment Preparation

A. Dataset and Hardware Environment

The data used in this study are high-resolution remote sensing images, and the data are provided by Jiangsu Bureau of Surveying and Mapping with a resolution of 0.3 meter.

Due to the large size of remote sensing images, it needs to be cut and segmented, and we selecte abundant images of different urban scenes as samples for training and testing. Finally, 58,640 images containing street trees are collected, and 45,000 of them are used as the training data set, 5,000 images are used as the valid test and the remaining 8,640 images are used as the test set. Meanwhile, the data set contain the street trees information of different varieties of trees and growth environments, which effectively improves the robustness of the model. During the period of data annotation, there would be only tree class labeled.

This experiment is conducted on windows10 operating system; memory size is 32G; GPU: NVIDIA GeForce 3070; the learning frameworks are tensorflow-gpu 1.13.1, keras 2.1.5, cuda10.0 and cudnn 7.4.1.5.

B. Evaluation Index

When evaluating the detection effect of the model, Precision, Recall and F1 of the model usually appear as vital indicators. The calculation Equations are Equation (1)- (3). Precision is based on the predicted results, showing the proportion of the number of correct samples in the total number of samples; Recall is the proportion of positive cases that can be correctly predicted.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{3}$$

TP stands for the number that was a street tree and correctly detected; FP stands for the number that is not a street tree, but is wrongly identified as a street tree; FN stands for the number which the street tree is not identified or wrongly identified. In the evaluation, the higher Precision and Recall, the closer the better detection effect would be. However, there is a contradictory relationship between the two, the most common way is by calculating F1 score.

Mean Intersection Over Union (MIoU): A standard measure for semantic segmentation. It calculates the parallel and cross-ratio of two sets. In semantic segmentation, these two sets are ground truth and predicted segmentation respectively. The calculation is in Equation (4).

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{TP+FP+FN} \tag{4}$$

Average precision (AP) is also a popular metric in measuring the accuracy of semantic segmentation like FCN, Unet, etc. AP computes the average Precision for Recall over 0 to 1. Mathematically, AP is defined in (5), where p stands for Precision and r stands for Recall.

$$AP = \int_0^1 p(r)dr \tag{5}$$

IV. Experiment and Discussion

A. Street Trees Detection and Contour Extraction

After the dataset trained by YOLOv4 network, 8,640 images in the test set are tested, with a total of 13,040 street trees.TP value is 11918, FP value is 639, FN value is 483. According to Eq. (1)-(3), the Precision, Recall and F1 score of the experiment can be calculated as 94.91%, 96.11% and 95.51%, respectively, showing that this model is strong. Here, Precision and Recall are consistent with the calculation methods of YOLOv4 evaluation indexes. We only labeled a tree class, so mAP is equal to AP and the MIoU is the same as IoU, and calculated by Unet is 98.25%, and AP is 99.29%.

Part of the detection results by YOLOv4 network and the segmentation effect of Unet network are shown in Fig.4. It is obvious that there exist omissions in the object detection task of street trees and the street trees can be basically segmented after semantic segmentation algorithm.

B. Integration of Street Trees Detection and Contour Extraction in Remote Sensing Images

The integrated processing model of street trees detection and contour extraction in remote sensing images is composed of two sub-models. The specific process is as follows: (1) the street trees are detected in remote sensing images through YOLOv4, and the coordinate information of the bounding box and confidence coefficient are obtained;(2) the images are sent to the Unet to classify the street trees pixel by pixel;(3) the detected result based on the YOLOv4 is fused with the corresponding region on the segmentation result to obtain the effect of street trees detection and contour extraction. The processing flow and effect is shown in Fig.5.

C. Estimation of Street Trees Coverage Ratio in Urban Scenes

Two different road scenarios were selected from the test set for discussion, including different varieties of trees and light environments. The test results are shown in Fig.6: straight road scene detection results; Fig.7: small roadside parking lot detection results, the left side of each image shows the result based on YOLOV4; and the right side shows the result of the proposed method.

The Equation (6) is to calculate the coverage ratio of street trees [17], where the coverage ratio of street trees is denoted as P, the total area of the vertical projection of street trees is denoted as S, and the total area of urban land is denoted as St. In the process of street trees detection, the information of the detection bounding box can be output together with the detection confidence coefficient. Here, S is acquired

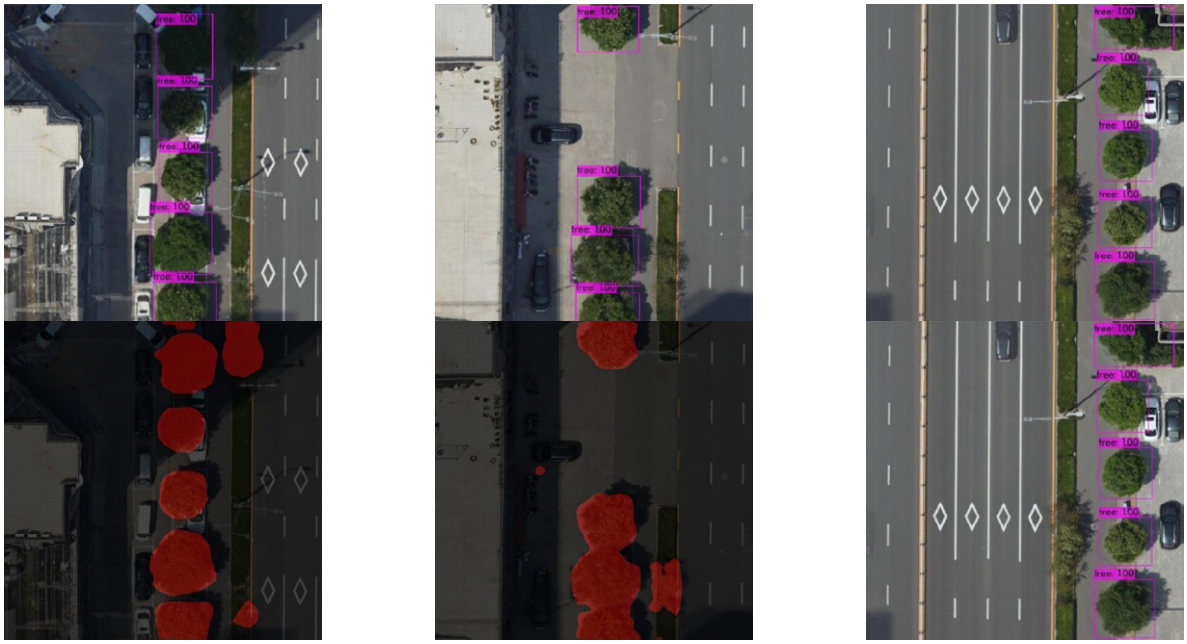

Fig. 4. (a) Results of street trees detection based on YOLOv4; (b) Results of street trees segmentation based on Unet.

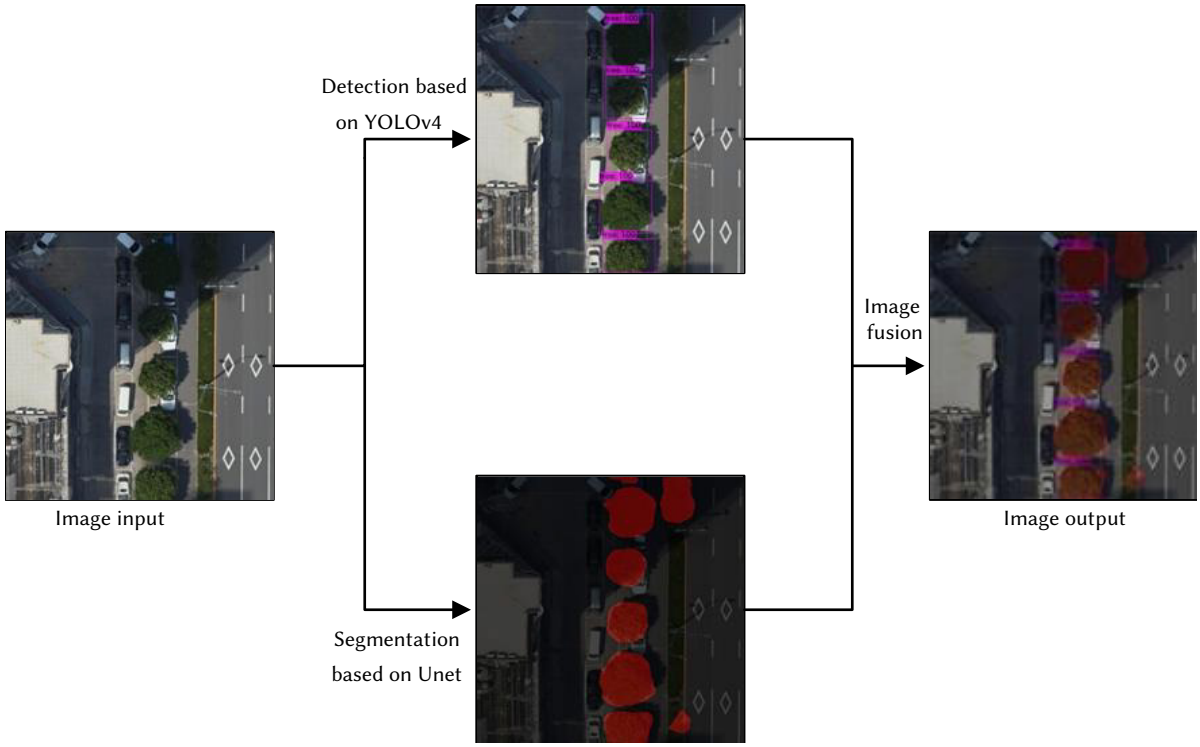

Fig. 5. Schematic diagram of integrated processing of street trees detection and contour extraction.

by the product of the obtained height and length. When the contour of the street trees is extracted, the size and area of the obtained contour pixel are calculated as the desired S, and St is the actual area of the detected remote sensing image.

$$P = \frac{S}{St} \tag{6}$$

Through Equation (6), we can get the street trees coverage ratio of these two scenes, we named the straight road scene and small roadside parking lot scene as S1 and S2 respectively. S1 and S2 based on YOLOv4 are 13.81% and 16.3% respectively, S1 and S2 based on the integrated processing model are 8.08% and 9.47% respectively. It is easy to see that the estimated value of urban street trees coverage ratio obtained by the proposed algorithm is close to the actual value.

D. Discussion

As can be seen from the above, object detection of street trees through YOLOv4 is in good condition, but there exist some cases that may lead to exceptional detection: when the illumination is uneven, it cannot be detected; when the street trees grow densely, the complete information of trees cannot be well detected; meanwhile, when the shape and outline of street trees have a visual difference, it will lead to omitting trees to be detected. However, the integrated method of street trees detection and contour extraction in remote sensing images can

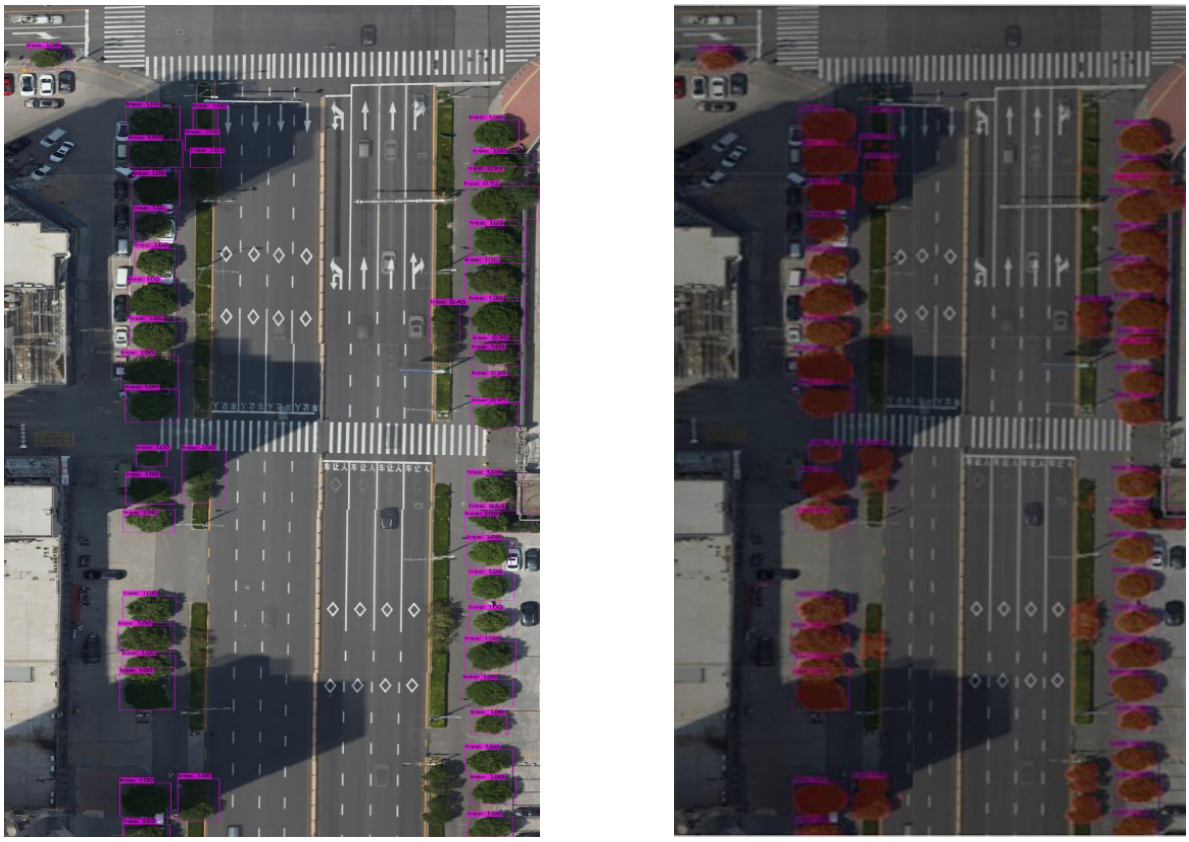

Fig. 6. Detection results of straight road scene.

Fig. 7. Scene detection results of small roadside parking lot.

solve these problems. For comparison, we used the Mask RCNN [18] to train the same data set. The following are the comparative results of each network model on the street trees, as is shown in Fig. 8.

As can be seen from Fig.8(a), street trees or part of street trees that are not detected in the YOLOv4 are marked with a red oval shape, including the following situations: street trees that cannot be detected in areas with weak illumination; where the street trees grow densely, the complete trees cannot be detected; meanwhile, when the shape and outline of the street tree have no obvious visual characteristics, there will be the situation of missing detection. All of the above will cause errors in coverage ratio calculation. As can be seen from Fig.8 (b), the green oval is used to mark the improvement and existing problems by the Mask RCNN. Undoubtedly, part of the above problems can be solved, there still exist some issues to settle. The street trees will be detected where the illumination is insufficient and there is also a partial missing of street trees in pixel detection. When this method was used for the identification and coverage ratio estimation of street trees, many issues will occur. For example, in the figure on the far left, there is a condition of repeated detection, so the coverage ratio of street trees will be bigger than the real value. In the other two images, the same problem appeares in the detection results of the YOLOv4, that is, in the detection process, there will be missed detection of street trees, the detection information of street trees is incomplete, and when the shape and outline of street trees has no obvious visual characteristics, there will lead to miss detection of street trees. The above cases will cause errors in the estimation of the street trees coverage ratio. As can be seen in Fig.8 (c), these problems can be totally solved through the integrated processing of detection and contour extraction model of the street trees. The yellow oval is used to mark the improvement and optimization of the information of the street trees can be obtained

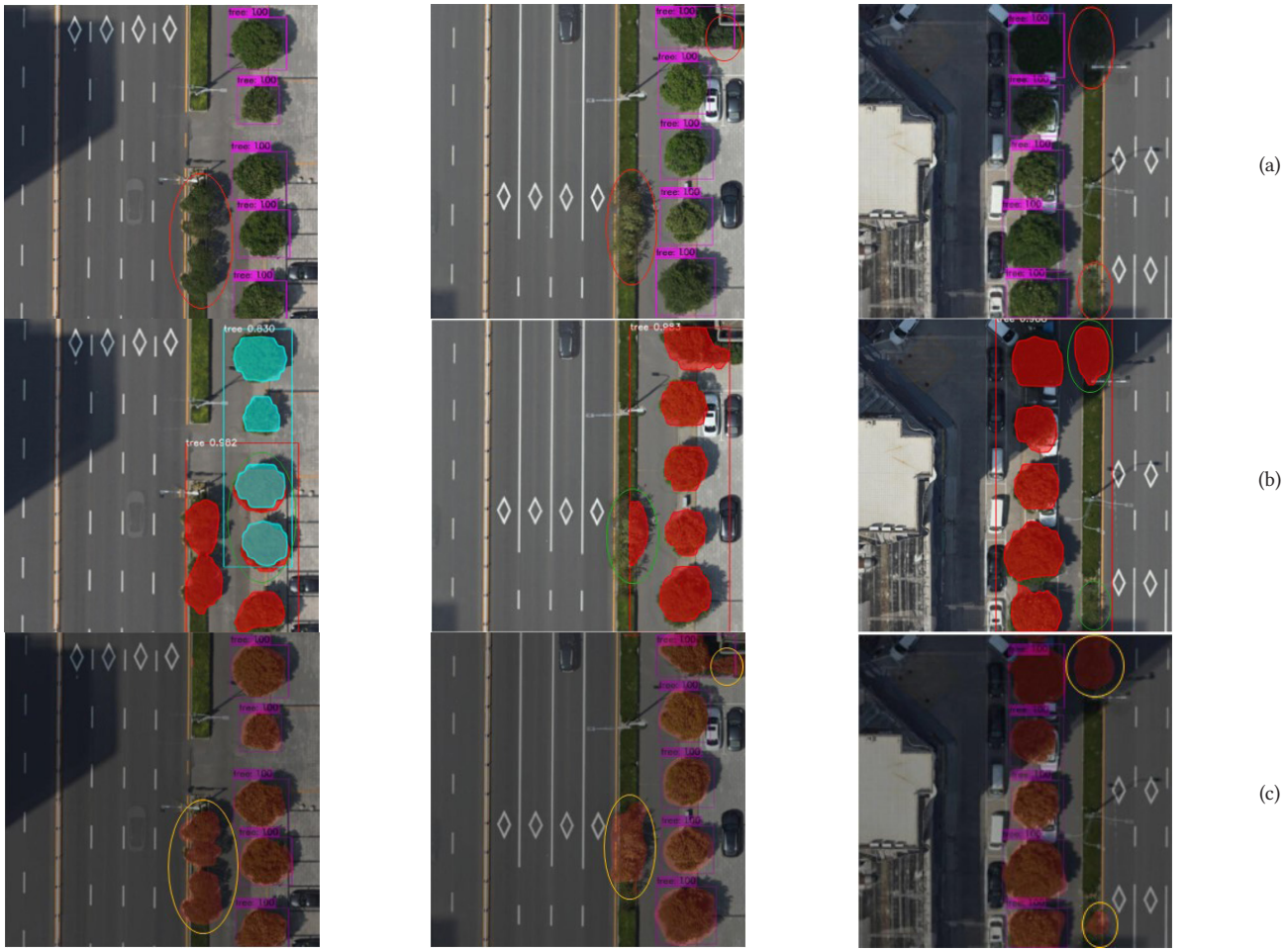

Fig. 8. (a) Results of Street tree detection based on YOLOv4;(b) Results of Street tree detection based on Mask RCNN;(c) The integrated detection results of street tree detection and contour extraction

through the proposed method. When calculating the covered area of the street trees, it was closer to the real value, which can improve the estimation accuracy of the covered area of the street trees in local scenes. In this case, the coverage ratio of the street trees obtained is closer to the real coverage ratio of the street trees themselves, so that the estimated coverage ratio of the street trees in local city scenes is closer to the real value.

E. Comparative Experiments With Other Mainstream Algorithms

The configuration information of the training platform remains unchanged. YOLOv3, YOLOv4, Mask RCNN and the proposed model were used for training and testing on the same data set. Comparison of mAP of different models is shown in Table I.

In the process of object detection, the mAP of the proposed model is superior to other mentioned models, indicating that the prediction accuracy of the proposed model is greatly improved compared with the method in the Table I.

TABLE I. Comparison of Model Performance

Detection model	mAP(%)
YOLOv3	81.72
YOLOv4	89.21
Mask RCNN	88.86
Proposed	99.29

F. Experiments Under VHR-10 Dataset

In order to verify the robustness of the proposed model, we also trained and tested the VHR-10 [19] [20] [21] dataset in different models. Table II listed the AP of five categories for the four methods (YOLOv3, YOLOv4, Mask RCNN, Proposed). We can find that AP values of these categories were significantly improved by detection of the proposed method, and the model was robust.

TABLE II. AP (%) of 5 Categories Targets of VHR-10 Dataset

Category	method			
	YOLOv3	YOLOv4	Mask RCNN	Proposed
Airplane	90.89	90.86	89.56	97.48
Baseball diamond	96.59	97.20	96.72	98.86
Tennis court	90.23	90.35	90.00	92.31
Ground track field	99.40	99.38	99.35	99.72
Bridge	86.78	87.32	87.35	88.85

V. Conclusion

In this study, we used the object detection method based on the YOLOv4 to carry out object detection on street trees and obtain the bounding box information and confidence coefficient at the same time, and calculated the coverage ratio of the street trees in urban scenes. Due to the non-tree-side part in bounding box or undetected trees, this method exits some errors. In order to solve these problems, we put forward the integration of remote sensing image object detection and contour extraction of street trees along the model to improve the accuracy of the coverage ratio.

The proposed model has the following three advantages when used to calculate the coverage ratio of street trees. Firstly, since there was only one tree class, a great loss value can be obtained by only iterating about 100 times when training under Unet network. Meanwhile, MIoU and AP are 98.25% and 99.29% respectively, compared with the mentioned models, the result is better. Then, b because there is no limit to the image size of the network, you can input any size images. Finally, through the proposed method, we can solve the problems occurring in street trees detection based on YOLOv4 and Mask RCNN, such as there exist the non-tree-side part in the bounding box and undetected trees while detecting, we can further obtain all information of street trees, and calculate the coverage ratio which is much closer to the real coverage ratio of trees.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62102184), in part by the Natural Science Foundation of Jiangsu Province (Grant No. BK20200784), in part by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (Grant No. 19KJB520010), in part by China Postdoctoral Science Foundation (Grant No. 2019M661852) and in party by the National Key Research and Development Program of China (2019YFD1100404).

References

[1] B.T. Wang, W.J. Wang, S.H. Cui, Y.Z. Pan, and J.H. Zhang, "Research on the methods of urban ecological environmental quality assessment," in Acta Ecologica Sinica, vol. 29, no. 3, pp. 1068-1073,2009.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," in NATURE, vol. 521, no. 7553, pp. 436-444, 2015.

[3] J. Kuruvilla, D. Sukumaran, A. Sankar, and S.P. Joy, "A Review on Image Processing and Image Segmentation," in International Conference on Data Mining and Advanced Computing, Kadayiruppu, INDIA, 2016, pp. 198-203.

[4] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in COMMUNICATIONS OF THE ACM, vol. 60, no. 6, pp. 84-90, 2017.

[5] Y. Long, "New ideas for urban research and planning under the new data environment of street urbanism," Time Architecture, vol. 02, pp. 128-132, 2016.

[6] X. Li, C. Zhang, W. Li., "Assessing street-level urban greenery using Google Street View and a modified green view index," in Urban Forestry & Urban Greening, vol. 14, no. 3, pp. 675-685,2015.

[7] I. Seiferling, N. Naik, C. Ratti, 2 "Green Streets-Quantifying and mapping urban trees with street-level imagery and computer vision," in Landscape and Urban Planning, vol. 165, pp. 93-101, 2017.

[8] L. Yan, Q. Xu, Y. Liu, "Vegetation extraction from remote Sensing Images based on attention network," in Surveying and mapping geographic information, vol. 46, no. S1, pp. 44-48, 2021.

[9] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, "Yolov4: Optimal speed and accuracy of object detection,"2020. Accessed: Apr. 23, 2020. [Online]. Available: https://arxiv.org/abs/2004.10934, doi: arXiv:2004.10934.

[10] N. Abbas, F.Q. Yu, "A Comprehensive Analysis of the End-to-End Delay for Wireless Multimedia Sensor Networks" in Journal of Electrical Engineering & Technology, vol. 13, no. 6, pp. 2456-2467,2018.

[11] P.F. Guo, X.Z. Wang, Y.S. Han, "The Enhanced Genetic Algorithms for the Optimization Design," in 2010 3RD International Conference on Biomedical Engineering and Informatics, pp. 2990-2994, 2010.

[12] K. He, X. Zhang, S. Ren, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 2015.

[13] S. Liu, L. Qi, H.F. Qin, J.P. Shi, J.Y. Jia, "Path Aggregation Network for Instance Segmentation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8759-8768,2018.

[14] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement," 2018. Accessed: Apr. 8, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767, doi: arXiv:1804.02767.

[15] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp. 234-241, 2015.

[16] J. Long, E. Shelhamer, T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, 2015.

[17] L.W. Yang, Z.X. Zeng, and C.J. Zhao, " Study on Statistical Technique of Urban Roadway Tree Greening Coverage," in Chinese Landscape Architecture, vol. 04, pp. 38-39, 1996.

[18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 386–397,2018.

[19] G. Cheng, J.W. Han, "A survey on object detection in optical remote sensing images," Journal of Photogrammetry and Remote Sensing, vol. 117, pp. 11-28, 2016.

[20] G. Cheng, P. Zhou and J. Han, "Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 12, pp. 7405-7415, 2016.

[21] Cheng G, Han J, Zhou P, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors" in ISPRS Journal of Photogrammetry and Remote Sensing, vol. 98, pp. 119-132,2014.

Wen Han

Wen Han received his B.S. degree in software engineering in 2018 from Yangtze University, Jing Zhou, China. Now, he is studying for the master degree at the College of Information Technology in Nanjing Forestry University. His research interests include computer vision and deep learning. He has published several papers and patents in domestic and international journals. He has two years of experience working in a fortune 500 company and is currently an intern in a research institute.

Lei Cao

Lei Cao received his bachelor degree in Communication Engineering in 2021 from Information College Huaibei Normal University, China. Now, He is studying for his master degree at the College of Information Science and Technology from Nanjing Forestry University. His research interests include Computer Vision and Image Processing.

Sheng Xu

Sheng Xu received his Ph.D degree from the University of Calgary in 2018, He joined the College of Information Science and Technology at Nanjing Forestry University. At present, he has published more than 40 papers on international and domestic journal conferences. In recent five years, he has published more than 10 SCI papers on IEEE Trans. TPAMI, IEEE Trans. TIST, IEEE Trans. His research interests include Three-dimensional spatial information processing, point cloud data analysis and computer vision.