SPRING POINT DETECTION OF HIGH RESOLUTION IMAGE BASED ON YOLOv3

Jiao Wu¹, Zhijun Zhang², Guopeng Huang³, Guorui Ma^{4,*}

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University - 1250230225@qq.com

² Comprehensive Survey Center for Natural Resources, China Geological Survey, Xining, Qinghai, China - zhangzhijun-0001@163.com

³ The Second Military Representative Office of the Rocket Army Equipment Department in Wuhan - 51679502@qq.com

⁴ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University -1366406@qq.com

Commission II, WG II/5

KEY WORDS: Spring point, YOLOv3, Target detection, Xinjiang region, Network framework, The Belt and Road Initiatives;

ABSTRACT:

The Xinjiang region of China is a vast and sparsely populated area with complex topography, surrounded by basins and mountains, and its geomorphological features and water circulation process make the traditional spring water resource acquisition time-consuming, labor-consuming and inaccurate. Remote Sensing Technology has the advantages of large scale, periodicity, timeliness and comprehensiveness in target detection. In order to realize the artificial intelligence detection of springs in Xinjiang, this paper presents a method of detecting springs in remote sensing image based on the YOLOV3 network framework, based on the data set of 512 * 512 by using 0.8m remote sensing image annotation, a model of recognition of spring point based on Yolov3 network is constructed and trained. The results show that the map of spring point is 0.973, which is the basis of monitoring and protecting the natural environment in the Belt and Road Initiatives.

1. INTRODUCTION

Remote sensing technology has the advantages of short cycle time, low cost, large coverage area, and convenient access. It can provide real-time, accurate, and large-scale surface information, and has been applied in a wide range of fields (Huang et al. 2018), (Kaku et al. 2019) and (Lassalle et al. 2020). Among them, target detection based on remote sensing technology can determine the type and location of objects at the same time, while reducing the use of human resources, which has important practical significance.

Spring point is one of the geological resource elements that need to be investigated in Xinjiang Uygur Autonomous Region of China, and it is also one of the basis for monitoring and protecting the natural environment in the Belt and Road Initiatives. Because of the large area and complex terrain in Xinjiang, it is very difficult to find springs, so the efficient and accurate detection of springs has become an essential part of the regional development planning and management of geological resources.

At present, there are two methods to detect the springs, which are the traditional manual visual image, field verification and deep learning object detection. Traditional manual interpretation of images generally requires observation of a composite model of color, texture, shape, etc. (Lee et al. 2019), (Liu et al. 2018), (Sun et al. 2019), (Xu et al. 2018) and (Yan et al. 2018). It has a high dependence on image pre-processing and feature selection, especially on the unrecognized spring point, which will be largely missed. Therefore, the universality of this method

is poor, and the field survey is difficult to promote in the field geological exploration. With the rise of deep learning, scholars tend to use deep learning to complete the target detection, which has better precision. The main methods can be divided into two categories. One type is a detection algorithm based on regional target recommendation. First, regional recommendations are used to generate candidate targets, and then convolutional neural network processing is used. But it cannot detect the target position and the real-time performance is not enough. Representative algorithms include RCNN (Girshick et al. 2014), Fast RCNN (Girshick. 2015), Faster RCNN (Ren et al. 2015), Mask RCNN (He et al. 2017), etc. Another type of method is the target detection method based on regression method, which converts the target detection problem into regression problem processing and directly predicts the target position and category. Thus, in 2015, Joseph Redmon proposed Yolo (You Only Look Once) Algorithm based on convolutional neural network and candidate region generating algorithm. Redmon et al(2016) transforms target detection task into regression problem, and the detection speed has made a breakthrough progress. Compared with Yolo Algorithm, YOLOv3 integrates the advantages of Faster R-CNN, SSD and ResNet, and the speed and precision of target detection is more balanced by Redmon and Farhadi (2018).

At present, the Algorithm of Yolov3 has achieved excellent performance in many fields. Good real-time performance with an average recognition accuracy of 88.7% based on YOLOV3 for automatic location identification and diagnosis of external power insulation equipment like (Liu et al. 2020). Kong et al. An improved YOLOv3 Algorithm for real-time detection of Underwater Sonar, which guarantees both detection speed and feature extraction capability was proposed like (Kong et al. 2020). YOLOv3-based deep learning network was used to detect

^{*} Guorui Ma - 1366406@qq.com

key parts of cows in complex scenes with an average accuracy of 99.18% like (Jiang et al. 2019). Therefore, this paper establishes a spring point recognition model based on the Yolov3 Algorithm, in order to obtain the precision and efficiency, thus laying a technical foundation for large area precise recognition of spring point.

2. DATA AND MATHODS

2.1 Data acquisition

Because there are few high-resolution spring point detection and recognition methods based on deep learning, and there is no public data set or standard data set. To this end, first of all, we should do the collection and construction of data sets.

The data range of spring sample is $(72^{\circ}, 38^{\circ}), (73.5^{\circ}, 38)$ $^{\circ}$),(72 $^{\circ}$,37 $^{\circ}$),(73.5 $^{\circ}$,37 $^{\circ}$), we download 0.8 m highresolution remote sensing image of this area, all unified CGCS2000 coordinate system. We use the artificially verified vector spring point data provided to us by the Qinghai Geological Survey Brigade, and use the selfdeveloped program to center the position of the vector sample points to crop the 0.8-meter high-resolution image into blocks. The Spring Data and the 317 scene of highresolution remote sensing image (Table1) are superimposed, and the training sample set is cut and normalized to get the final positive sample set. The specific steps are: run cutting script file, all unified cut into 512x512 size image blocks, a total of 679 JPG format spring sample image.

Year	High-resolution images (tif)	Resolution	Vector spring point (shp)
2019	j43c0030011 J43C0030021 j43c0030031 J43C0040011 j43E00110017	0.8m	J43C003001Q J43C003002Q J43C003003Q J43C004001Q J43C004002Q J43C004003Q

Table 1. Data used in the experiment

First of all, we should do the data annotation. We use manual labeling. In order to ensure the validity of the data, the samples with obvious targets are marked. In the end, the spring point targets in 679 multi-spectral images were labeled with bounding boxes, the interface of the labeling system is shown in Figure 1, and data with labeling information was obtained. According to the ratio of 9:1, it is divided into training set and test set.



Figure 1. Data set annotation

Second, we need do the image enhancement. In order to improve the detection and recognition accuracy of the method, the training set is enhanced by the image enhancement method commonly used in the target detection field, including the operation processing of brightness and contrast.

2.2 Experimental method

The Yolo Algorithm proposed by Redmon et AL in 2016 transforms the target detection task into the regression problem, greatly speeds up the detection speed. On the basis of YOLOv2, YOLOv3 is proposed to maintain the detection speed of YOLOv2, and greatly improve the detection accuracy, especially in the detection and identification of small targets.

First, the original image is scaled to the size of 416x416, and the original image is divided into SxS cells. It detected on three different scales of feature maps such as 13x13, 26x26, 52x52. Each cell has three anchor boxes to predict three bounding boxes.

The convolutional neural network predicts four values for each bounding box on each cell, namely the target box (x, y), we take width W and height H as t_x , t_y , t_w , t_h , respectively. The target center is in the cell relative to the image left, the upper corner is offset $b_y c_x$, c_y , and the anchor box has height and width p_w , p_h , the modified bounding box is:

$$\mathbf{b}_{x} = \boldsymbol{\sigma}(t_{x}) + c_{x} \tag{1}$$

$$\mathbf{b}_{\mathbf{y}} = \sigma(t_{\mathbf{y}}) + c_{\mathbf{y}} \tag{2}$$

$$\mathbf{b}_{\mathrm{w}} = p_{\mathrm{w}} e^{t_{\mathrm{w}}} \tag{3}$$

$$\mathbf{b}_{\mathbf{h}} = p_{\mathbf{h}} e^{t_{\mathbf{h}}} \tag{4}$$

In the course of training, the sum of squares of errors is used as the loss function. If the true coordinates are t_w , the gradient can be obtained by minimizing the loss function. The gradient is the true coordinates minus the predicted coordinates: $t_w - t_h$.

Yolov3 used the newly designed Darknet-53 in the feature extraction phase. The residual connection is widely used to increase the depth of the network, and combined with the FPN network structure, the two characteristic graphs are sampled on the back of the network, and then the corresponding characteristic graphs are aggregated with the previous size of the network, and the predicted results are obtained by the convolution network. In the phase of target detection, Yolov3 firstly predicts the feature map of 13x13 by convolution, and gets the first detection result, and then the 26x26 feature map sampled from the 13x13 feature map is fused with the 26x26 feature map in the previous network to form a new feature map, the second detection result is obtained by multi-convolution, and then 52x52 feature map is obtained by fusing 26x26 feature map with the previous layer, and the third detection result is obtained by convolution (Figure 2). Finally, the final recognition result is obtained by non-maximum suppression of the three detection results.



Figure 2 YOLOv3 target detection framework

Yolov3 transforms the target detection task into the regression task of target region prediction and class prediction, and uses a single neural network to predict the object boundary and class probability directly, and realizes the end-to-end object detection (Yi et al. 2019). In contrast to other target detection algorithms (such as ssd, and Yolo with different final full connection layer, ssd directly use convolution to extract detection results from different feature maps.)

In the training process of spring detection, the flow chart of Yolov3 can be expressed as follows (Figure 3): First, the input image is divided into S*S cells, and 3 bounding boxes are predicted for each cell, using convolution neural network.Finally, the position of the bounding box, the object score and the class probability are predicted by using the repressor.



Figure 3. YOLOv3 mode l training process

3. EXPRIMENTS AND RESULT

3.1 Test Platform

The experiment was completed in Linux, with Ubuntu 16.04 operating system, CUDA 9.0, Cudnn7.0, Opencv4.1, Python3.6 and other third-party libraries installed to support the pytorch framework. The computer has 62.9 GB of Ram and runs on an Intel

Xeon(R) CPUE5-2687Wv4 @3.00GHz×8 processor. The Graphics Card uses NVIDIA Virtual Graphics Card GRID RTX8000-12Q, with the memory of 12G.

3.2 Data Set Construction

Firstly, Labeling is used to annotate 0.8m remote sensing image, which is stored as Pascal format XML file, and then XML file is converted into TXT file of <label, x, Y, W, H> format by format conversion script. Different data sets were made, From 679 images, and 245 original data sets were selected for better quality.

The network structure is based on Python's Pytorch library. During the training process, in order to prevent overfitting, data enhancement processing is added. We used a self-programmed program to complete the data enhancement process. After the image, clipping, contrast, brightness, noise and other data enhancement processing, finally got 1545 sample data for testing springs.



R90_Blur.jpg

R90_Brighter.jpg

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition)



R180_Blur.jpg R180_Brighter.jpg Figure 4. Partial results of data enhancement with a single JPG

The above is a part of the data enhancement result of a single image after processing brightness, shading, cropping, noise, and rotation by 90 degrees and 180 degrees. Each image sample collected by the experiment has undergone the above enhancement processing.

3.3 Model Training

3.3.1 Training method: In this experiment, a mini-batch stochastic gradient descent with momentum factor (momentum) was used to train the model for 1000 iterations. In order to avoid over fitting, the regularization coefficient was set to 0.0005, and the momentum factor was set.

At the same time, the training set and the test set are divided into 9:1.In order to test the validity of the expanded dataset, the original data and the expanded data were compared and analyzed. The implementation of Yolov3 is based on open source code (https://github.com/ultralytics/yolov3).

3.3.2 Model evaluation and criteria:

The target detection of the springs needs not only the precision, recall, F1- (F1-measure, F1) and we also take detection speed as the evaluation criteria, but also the average precision (mean average precision, map). The formula is as follows:

$$P=TP/(TP+FP)$$
(5)
$$R=TP/(TP+FN)$$
(6)

$$F1=2P \times R/(P+R)$$
(7)

$$map=APspring point$$
 (8)

TP is the true case, FP is the false positive case, FN is the false negative case, and APspring point is the single class precision of spring point.

3.4 Results Analysis

3.4.1 Model evaluation: In the process of model training, because the parameters such as learning rate and the value of loss function are changing constantly, the quality of the model and the number of iterations is not completely linear, therefore, it is necessary to evaluate the accuracy and the convergence of a model. Only when the accuracy of the model is high and the convergence is good, the model can maintain robustness, robustness and stability when making predictions. In this article, Yolov3 performs a fountain-Loss test with a Loss curve as shown in the following figure:



Figure 5. Relationship between iteration number and loss function

As we can see in figure 5, with the number of iterations increasing, by the time of 1000 iterations, the loss has been reduced to about 0.6 and stabilized. Training 1000 epochs, map up to 0.973.



This contribution has been peer-reviewed.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020, 2020 XXIV ISPRS Congress (2020 edition)



A series of indicator graphs will be generated during the training of experimental. From the above picture, the indicators have been mentioned in section 3.3.2, the value of GIOU, Objectness, Val GIOU, Val Objectness are gradually decreasing and steadily approaching a certain value. We can see the precision gradually began to rise, there were fluctuations in the middle, and the last time it suddenly dropped, because the value of IOU threshold for non-maxium suppression was not set, in the experiment I set 0.5, so there will be no problem of the final dip in precision. F1 and accuracy have the same phenomenon and problems. The value of map and recall showed an upward trend and eventually stabilized. They can show a better spring point target detection result.

When the confidence threshold is higher, the accuracy of model detection is higher, but the recall rate is significantly reduced, indicating that all detection targets cannot be detected. When the confidence threshold is set lower, the recall rate increased but the accuracy of model detection decreased. In order to make the detection accuracy and recall rate of the model reach a good level at the same time, the confidence threshold is 0.5 and the map value is 0.973. Both accuracy and recall rate show better results. Its accuracy is 0.901, and recall is 0.983.

3.4.2 Test results: The experiment selected 375 pictures for prediction. The effect of the test is shown in the figure 7. The spring point detection frame and confidence interval on the test data of high-resolution remote sensing images can be clearly seen.





Figure 7. Test result of spring point recognition

In the experiments in this paper, map (mean Average Precision) is used to judge the effect of model recognition. map is a commonly used indicator to measure detection accuracy in target detection, referring to the average accuracy of multiple categories. AP refers to It is the average accuracy of a single category. It measures the quality of the model's recognition in a certain category. The map used in this article is the average of all APs. The larger the value of map, the higher the overall recognition accuracy of the model.

4. CONCLUSION

Compared with the traditional fountain recognition technology, the deep learning method saves a lot of manpower and material resources, in this paper, we use YOLOv3 Algorithm to identify spring point in Xinjiang region through high resolution remote sensing image data, its map reaches 0.973, but still need a lot of research work. The following work should include analysis and evaluation experiments for the effective estimation of the sample size for all regions in the proposed validation work and the final design of the validation scheme.

The accuracy of spring spot samples is very important for the experiment. The next step will continue to improve the samples and mine more suitable spring spot detection algorithms. The research work in this paper is of great significance for the investigation of the detection of geographic elements along the Belt and Road.

ACKNOWLEDGEMENTS

Acknowledgements of support for the project provided by the Comprehensive Survey Center for Natural Resources, China Geological Survey, Xining, Qinghai, China.

REFERENCES

Girshick R., 2015: Fast R-CNN. International Conference on Compute Vision, 1440-1448. doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., & Malik, J., 2014: Rich feature hierarchies for accurate object detection and semantic segmentation. *computer vision and pattern recognition, Columbus,* 580–587. doi.org/ 10.1109/CVPR.2014.81.

He, K., Gkioxari, G., Dollar, P., & Girshick, R., 2017: Mask R-CNN. *international conference on computer vision*, 2017:2980–2988. doi/org/ 10.1109/iccv.2017.322.

Huang, Y., Chen, Z., Tao, Y. U., Huang, X., & Xingfa, G. U., 2018: Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, 17(9): 1915-1931. doi/org/ 10.1016/S2095-3119(17)61859-8.

Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., & He, D., 2019: FLYOLOv3 deep learning for key parts of dairy cow body detection. *Computers and Electronics in Agriculture*. doi/org/ 10.1016/j.compag.2019.104982.

Kaku K., 2019: Satellite remote sensing for disaster management support: A holistic and staged approach based on case studies in Sentinel Asia. *International Journal of Disaster Risk Reduction*, 417-432. doi/org/10.1016/j.ijdrr.2018.09.015.

Lassalle, G., Fabre, S., Credoz, A., Dubucq, D., & Elger, A., 2020: Monitoring oil contamination in vegetated areas with optical remote sensing: a comprehensive review. *Journal of Hazardous Materials*. doi/org/10.1016/j.jhazmat.2020.122427.

Lee, M., Park, K.-A., Chae, J., Park, J.-E., Lee, J.-S., & Lee, J.-H., 2019: Red tide detection using deep learning and high-spatial resolution optical satellite imagery. *International Journal of Remote Sensing*, 1-23. doi/org/ 10.1080/01431161.2019.1706011.

Liu, J., Liu, S., Liu, W., Zhou, S., Zhu, S., & Zhang, Z.-J., 2018: Persymmetric adaptive detection of distributed targets in compound-Gaussian sea clutter with Gamma texture. *Signal Processing*, 152, 340-349. doi/org/ 10.1016/j.sigpro.2018.06.006.

Ren, S., He, K., Girshick, R., & Sun, J., 2015: Faster R-CNN: towards real-time object detection with region proposal networks. *neural information processing systems*,:91–99.

Redmon, J., Divvala, S. K., Girshick, R., & Farhadi, A., 2016: You Only Look Once: Unified, Real-Time Object Detection.:*Computer Vision and Pattern Recognition.* doi/org/ 10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2018: YOLOv3: An Incremental Improvement. *Computer Vision and Pattern Recognition*.

Sun, Q., Li, L., Xin, Y., 2019: Research on the multiscale low rank method and its optimal parameters selection strategy for infrared small target detection. *Optik.* doi/org/ 10.1016/j.ijleo.2019.06.020.

Xu, S., Shi, X., Xue, J., Shui, P., 2018: Adaptive subspace detection of range-spread target in compound Gaussian clutter with inverse Gaussian texture. *Digital Signal Processing*, 81, 79-89. doi/org/10.1016/j.dsp.2018.07.002.

Yan, J., Du, J., Young, Y., Chatwin, C., Young, R., Birch, P., 2018: Real-time unmanned aerial vehicle tracking of fast moving small target on ground. *Journal* of *Electronic Imaging*, 27(5). doi/org/10.1117/1.JEL27.5.053010.

Yi, Z., Yongliang, S., Jun, Z., 2019: An improved tinyyolov3 pedestrian detection algorithm. *Optik*,17-23. doi/org/ 10.1016/j.ijleo.2019.02.038.