Implementation of Detection System of Grassland Degradation Indicator Grass Species Based on YOLOv3-SPP Algorithm

Shuo Dong 1, Yihan Ma 1 and Chunmei Li 1, *
1Department of Computer Technology and Application, Qinghai University, Xining, Qinghai Province, CHXX0236, China
*Corresponding author’s e-mail: li_chm0422@sina.com

Abstract. Due to climate change and human factors, grasslands in the Sanjiangyuan area have degraded. At present, most grassland workers use manual methods such as visual inspection, measurement and remote sensing technology or neural networks to conduct macro evaluation of grassland. However, the emergence of grassland degradation indicator grass is an important sign for grassland degradation. Therefore, it is simpler and more convenient to provide early warning of grassland degradation through the detection technology of degradation indicator grass species. In this paper, the degradation indicator grass species Stellera chamaejasme was used as an example, and the YOLOv3-SPP algorithm was used to detect the degradation indicator grass species. First of all, collect and process data on the spot, and establish a data set of Stellera chamaejasme for deep learning; secondly, use YOLOv3-SPP algorithm to train and test the data set. By continuously improving the quality of the data set, the accuracy of model detection is improved. Then through the test of the verification set, the accuracy rate reaches 95% and the recall rate reaches 98%. It is proved that the model can be used to detect grassland degradation indicator grass species under complex grassland background. Finally, the detection system of Stellera chamaejasme with uploading and detecting functions is realized.

1. Introduction
In recent years, due to climate change and overgrazing, the large-scale grassland degradation has caused various ecological problems, such as desertification and sandstorms [1]. According to the survey of grassland resources in Qinghai Province in 2012, there are 3131.04×104hm2 of degraded grassland in Qinghai Province, accounting for 81.03% of the available area of natural grassland in the province. The area of soil erosion accounts for 46% of the total land area of the province. Due to the degradation of grasslands, the diversity of grasslands has rapidly decreased. The threatened species in the province have accounted for 10%-20% of the total, which is 5 percentage points higher than the world average. Therefore, grassland degradation assessment is essential to prevent global land desertification and sandstorms. The emergence of grassland degradation indicator grass means that the trend of grassland degradation appears[2], and many countries have successfully used specific grass species as indicators for evaluating grassland degradation [3-4]. In a large number of grassland pictures, it is a good solution to use the detection of degradation indicator grass species to judge the degree of grassland degradation.

Studies on the degraded grasslands of the Qinghai-Tibet Plateau indicate that as the grasslands deteriorate, the coverage of Stellera chamaejasme flowers gradually increases. Therefore, Stellera chamaejasme is regarded as one of the important indicator grass species for grassland degradation.
The traditional grassland degradation assessment method relies heavily on human eyes and physical labor, which inevitably leads to subjective results and high labor costs. In contrast, the image target detection algorithm based on deep learning can detect the grass degradation indicator species well and can be used as the basis for the next step of grass degradation evaluation. In this paper, under the complex background of diverse grass species on the grassland, the degradation indicator grass species Stellera chamaejasme was taken as an example to study the detection methods of the degradation indicator species on the grassland.

Image-based plant detection is an important solution for plant taxonomy [5-7], however, this experiment is more challenging than these existing works. In these studies, plant recognition based on deep learning is more regarded as an image classification task. However, in addition to distinguishing degradation detection grass species in this experiment, it is necessary to specifically detect the position of degradation detection grass species relative to the image. In addition, these algorithms only detect this images which have the background and identifiable plants, while it is difficult to detect for the grassland where the grass species is located because it has a complex and similar background.

In the paper, the following key tasks are mainly done. First of all, a large amount of Stellera chamaejasme video and picture data were collected in the field in the Sanjiangyuan area. In the later stage, the video is decomposed to delete unclear and useless photos. Manually label photos with uniform size standards and flowers, and finally generate image data containing more than 20,000 pictures, creating a deep learning data set. Secondly, use YOLOv3-SPP algorithm and image cropping technology to crop the picture, in order to expand the size of the data set and improve the accuracy of detection. Finally, the detection accuracy rate was increased to 95% which proves the correctness of the experimental method for the detection of Stellera chamaejasme under complex grassland background. Finally, the detection system of Stellera chamaejasme with uploading and detecting functions is realized.

2. Related work

Materials Early target detection methods were based on sliding windows, using manually extracted features and classifiers on dense image grids to find targets. With the development of deep neural networks, the best results of target detection tasks are quickly surpassed by CNN-based detectors, and these detectors can be divided into single-stage and two-stage categories. The two-stage target detection algorithm consists of two parts. The first stage (such as heuristic search [8], EdgeBoxes [9], DeepMask [10-11], and RPN [12]) generates a sparse set of candidate targets. The second stage determines the accuracy of the every target detection area and the corresponding category label by the convolutional network. The biggest problem of the two-stage target detection algorithm is the low detection efficiency. In order to solve this problem, the single-stage target detection algorithm came into being. This method can be trained end-to-end and can go directly from the original pixels to different categories. While YOLO [13] achieves a faster detection speed, it uses a feed-forward convolutional neural network to directly predict the category and location. After this, YOLOv2[14] and YOLOv3[15] have been proposed one after another, further improving YOLO in several aspects, such as adding batch regularization on all convolutional layers, using high-resolution classifiers and convolutional layers instead of fully connected to predict the bounding box, etc.

3. The detection method of degradation indicator grass

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

3.1. Data source

Since there is no publicly available degraded grass species detection data set, this experiment builds a grassland degradation indicator grass species data set with videos and pictures taken on the ground. The specific statistics are shown in Table 1.
### Table 1. Data set acquisition information statistics

| Shooting Equipment   | Picture | Video | Total |
|---------------------|---------|-------|-------|
| Mobile Devices 1    | 50      | 71    | 121   |
| Mobile Devices 2    | 207     | 212   | 419   |
| Mobile Devices 3    | 3440    | 84    | 3524  |
| Mobile Devices 4    | 0       | 77    | 77    |
| UAV                 | 0       | 4     | 4     |
| **Total**           | 3697    | 448   | 4145  |

Using a third-party library of Python, the 448 video clips are processed, and every 25 frames are processed to form a piece of picture data. A total of 20027 picture files are generated.

3.2. **Preprocessing of picture data**

First unify the format of the data set, and then use the LabelImg image annotation tool to annotate the picture. Since this experiment only needs to detect the selected Stellera chamaejasme flowers, there is only one category in the annotation file of this article, which is named flowers. During the labeling process, check the pictures one by one, remove the blurry pictures and the pictures which lost due to problems such as file loss in the storage process, and then keep the clear pictures. Finally, the data set used in this experiment is obtained.

3.3. **Labeling strategy**

The data set of this experiment contains individual flowers, multiple overlapping flowers, or a cluster of flowers with strong aggregation (as shown in Figure 1). The labeling strategy is that when the Stellera chamaejasme flowers are dispersed, each flower All need to be marked. When the Stellera chamaejasme flowers cluster together and overlap or the target is small, mark a cluster of flowers (as shown in Figure 2).

#### Figure 1. Different forms of Stellera chamaejasme flowers

#### Figure 2. Corresponding labeling strategies for different forms of Stellera chamaejasme flowers

3.4. **Image cropping**

In order to obtain clear videos and pictures, the pixels set by the shooting device are high, and the pixels of the data set pictures are 3000*4000. If a large pixel image is input into the network model, the network model will be overloaded, and after reshaping, the objects in the image, especially small
objects, will lose their coordinate positions, resulting in inaccurate target detection. First read the size of the original image, set the cropped image to a ratio of 1:5 in the original image, the width step is width/20, and the length step is height/20. During the cropping process, for a case where a candidate frame is cropped, if the length and width of the remaining target frame is greater than or equal to 1/2 of the original, the target frame remains, otherwise it is discarded. If there is no label in the xml file corresponding to the cut image, it will not be saved, otherwise the cut image and its corresponding xml file will be saved. According to this method, crop 500 original picture data sets to obtain 106141 picture data sets of 600*800 specifications. The flow chart is shown in Figure 3.

![Image cropping flowchart](image-url)

**Figure 3. Image cropping flowchart**

**4. Model introduction**

YOLOv3 - SPP is based on YOLOv3 network increased the SPP module, and the module referenced by the ideas of the space of the pyramid through SPP module implements the global features and local features. The nuclear size of the biggest pooling in SPP module is as far as possible close to or equal to a drawing of the characteristic of pooling. After the integration of global features and local features, the size of the characteristics of the figure enriches the figure characteristics of the power of expression which is advantageous to the differences in the image to be detected target size. It has a great improvement on the accuracy of detection. The structure diagram is shown in Figure 4 below.

![YOLOv3-SPP model structure](image-url)

**Figure 4. Yolov3-SPP model structure**

**5. Model training**

According to the ratio of 9:1, the data set is divided into training set and test set. That is, the training set contains 18000 pictures and the test set contains 2000 pictures. In order to verify that the image clipping improves the accuracy of the inspection, the network model is trained by using the image clipped image data set and the clipped image data set respectively. The model after training is recorded as model 1 and model 2 respectively.
The operating system used in this experiment is Ubuntu 18.04, the hardware configuration is GeForce GTX 1080Ti GPU (16G memory), the integrated development tool used is PyCharm, Python version is 3.6, VNC Viewer remote login software, the network development framework used is Pytorch and YOLOv3-SPP target detection algorithm. The number of detection categories is set to the number of categories to be detected, which is 1. The learning rate is set to 0.001, the batch size of the training network model is 16, and the number of training rounds is set to 100. The training set is input to the network for training, and training parameters, models and training logs are saved during the training process.

6. Experimental results
The data sets before and after cropping were imported into the Yolov3-SPP network structure for in-depth training, and the verification set containing 2000 small pixel images after cropping was used to evaluate model 1 and Model 2 respectively.

6.1. Expand data scale
After data cropping, the original 3000*4000 pixel data is converted into a 600*800 pixel data set, which not only makes the pixels of the picture more suitable for network training, but also increases the size of the data set.

6.2. Comparison of test results
Figure 5, Figure 6 and Figure 7 show the comparison between the image to be detected and the detection effect between Model 1 and Model 2. Comparing Figure 6 and Figure 7, it can be seen that the detection effect of Figure 7 is better. This shows that the data set cropping technology improves the accuracy of the detection work and greatly improves the target's IoU value. It shows that the prediction box is closer to the real box. Figure 8 shows the change curve of the IoU value of the detection results of Model 1 and Model 2.

Figure 5. The image to be detected
6.3. Comparison of model evaluation indicator
The image clipping technology has greatly improved the evaluation index. Accuracy increased from 56.7% before cropping to 95% after cropping, recall rate increased from 34.3% to 98.7%, mAP value increased from 25.9% to 97.9%, F1 increased from 42.7% of Model 1 to 96.8% of Model 2. The above data indicate that in terms of the accuracy and correct number of detection, model 2 has a great improvement compared with model 1. Table 2 shows the comparison of evaluation indexes.
Table 2. Comparison of evaluation indicators

| The model number | Evaluation indicators |
|------------------|-----------------------|
|                  | Accuracy  | Recall  | mAP    | F1      |
| Model 1          | 0.567     | 0.343   | 0.259  | 0.427   |
| Model 2          | 0.950     | 0.987   | 0.979  | 0.968   |

6.4. Verification set mAP comparison

During the training process, the change curve of the mAP of the validation set is shown in Figure 9. The mAP amplitude before cropping is large and the mAP value is low; while the mAP after cropping reaches 98% around 30 rounds, and after the training process, the value stabilized at 98%. This shows that the cropped data set is more suitable for network training.

Figure 9. Comparison of mAP curves

7. Implementation of the detection system

Use PyQt4 to build an application for real-time detection of Stellera chamaejasme, including three parts: graphical interface, file upload and trans-coding, and Stellera chamaejasme detection. The system environment is Windows 10, and the Python version is 3.6. Figure 10 is the initial interface of the system; Figure 11 is the upload interface. File upload is also implemented by calling the PyQt4 module. The suffix for opening files is specified as ".jpg" or ".png". Because PyQt4 has a poor reading effect on the jpg image format, for the selected file with the suffix ".jpg", convert it to PNG format through OpenCV and open it. The area of the image to be detected on the right is filled with the open file. As shown in Figure 12, it will be automatically updated after the detection of Stellera chamaejasme.

Figure 10. Initial interface of Stellera chamaejasme detection system

Figure 10. Initial interface of Stellera chamaejasme detection system
8. Conclusion
This article is based on the pytorch deep learning framework YOLOv3-SPP target detection algorithm in grassland degradation indicator grass species Stellera chamaejasme. In this paper, firstly, the data set of grassland degradation indicator Stellera chamaejasme was created, and the large resolution image was cropped to make it more suitable for yolov3-SPP algorithm, thus improving the detection accuracy, recall rate, mAP and F1. The comparison test before and after the picture separation proves that the cropping of the data set verifies the successful detection of Stellera chamaejasme flower under the complex background of grassland. Finally, the detection system of Stellera chamaejasme with uploading and detecting functions was realized.

Author Contributions:
Data curation, Yihan Ma; Methodology, Shuo Dong; Software, Shuo Dong; Supervision, Chunmei Li; Validation, Yihan Ma; Visualization, Shuo Dong; Writing – original draft, Shuo Dong; Writing – review & editing, Chunmei Li. All authors have read and agreed to the published version of the manuscript.

Funding
The research was not supported by external founds.

Acknowledgments:
Special thanks to the experimental hardware conditions provided by the Department of Computer Science and Application of Qinghai University.
Conflicts of Interest:
The authors declare no conflict of interest.

References
[1] Wang Z, Deng X, Song W, et al. What is the main cause of grassland degradation? A case study of grassland ecosystem service in the middle-south Inner Mongolia[J]. Catena, 2017, 150: 100-107.
[2] Le-Le C, Jian-Jun S, Yan-Long W, et al. Study on Different Degraded Degrees Grassland Community Structure Characteristics of the Alpine Area[J]. Acta Agrestia Sinica, 2016.
[3] Mansour K, Mutanga O, Everson T, et al. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 70: 56-65.
[4] Mansour K, Mutanga O, Adam E, et al. Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors[J]. Geocarto International, 2016, 31(5): 477-491.
[5] Sarah B, Rihab B A, Guillaume C, et al. Bark and Leaf Fusion Systems to Improve Automatic Tree Species Recognition[J]. Ecological Informatics, 2018, 46:57-73.
[6] Ameur R B, Valet L, Coquin D. Fusion system based on belief functions theory and approximated belief functions for tree species recognition[C]// Sixth International Conference on Image Processing Theory. IEEE, 2016.
[7] Goéau H, Bonnet P, Joly A. Plant identification in an open-world (lifeclef 2016)[C]. 2016.
[8] MoonHee Cho, TaeJeong Kim. Selective Search for locating Object Hypotheses by Using Graph Cut[J]. 2017.
[9] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[C]//European conference on computer vision. Springer, Cham, 2014: 391-405.
[10] Pinheiro P O, Collobert R, Dollár P. Learning to segment object candidates[C]//Advances in Neural Information Processing Systems. 2015: 1990-1998.
[11] Pinheiro P O, Lin T Y, Collobert R, et al. Learning to refine object segments[C]//European Conference on Computer Vision. Springer, Cham, 2016: 75-91.
[12] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
[13] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 777-788.
[14] Redmon J, Farhadi A. YOLO9000: better, faster,stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
[15] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. https://arxiv.org/pdf/1804.02767.pdf.[2019-04-24].