Research on The Method of Grass Mouse Hole Target Detection Based on Deep Learning

Jiaoxia Wan¹,², *, Dongfang Jian¹,², Dongzhao Yu³

¹Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010000
²Key Laboratory of perception technology and intelligent system of Inner Mongolia Autonomous Region

*Corresponding author: wanjiaoxia@imut.edu.cn

Abstract. Due to climate warming, increasing soil erosion and overloading of grassland grazing, grasslands around Taipusi Banner in Inner Mongolia have been degraded to varying degrees, causing widespread rodent damage. Aiming at the problem of rodent damage in the grassland surrounding Taipusi Banner, the UAV visible light image is the main research object, and the precise detection of grass mouse holes is the research goal. Computer vision technology is used to study grass mouse hole detection methods, in order to achieve rapid and accurate mouse detection. Hole function.

Keywords: Visible light image, deep learning, target detection, mouse hole.

1. Introduction

According to statistics, the grassland areas of Inner Mongolia have been threatened by rodents for years. In areas with severe rodent damage, due to the large amount of forage grass being eaten, the surface of the grassland is bare and soil erosion is not optimistic. This not only poses a threat to the grassland ecological environment, but also seriously affects the normal production and life of farmers and herdsmen. Accurate and rapid detection of grassland rodent holes lays the foundation for real-time and high-efficiency monitoring of grassland rodent damage, and is the basis for taking corresponding measures for grassland ecological protection. Therefore, it is a very meaningful thing to carry out research on grass hole detection.

Traditional rathole detection methods are based on manual investigations, that is, scientific researchers judge the spatial distribution and density of ratholes in the entire grassland by analyzing the spatial distribution and density of ratholes in multiple sample plots. Although the manual investigation method is simple and easy to implement, this method is time-consuming and labor-intensive. With the rapid development of satellites, scholars at home and abroad have introduced satellites to the detection of grassland rodent pests. However, due to the impact of satellite spatial resolution, most rodent pest studies are conducted by detecting and analyzing environmental factors such as vegetation and topography in grassland areas. Indirect detection of rodent damage. In recent years, the technology of light and small drones has gradually developed. Its centimeter-level high spatial resolution makes rodent detection no longer only satisfied with studying the ecological environment around the rodent area. How to quickly and accurately detect grass rodent holes has
become Problems that need to be resolved [1]. Aiming at the problem that the maximum likelihood method and the object-oriented method are not accurate in extracting the mouse hole [2], a target detection method based on deep learning is proposed.

2. data collection

2.1. Overview of the study area

Taipusi Banner is located at the southernmost tip of Xilin Gol League in Inner Mongolia. It is the closest flag to Beijing. Its geographic coordinates are between 114°51′~115°49′ east longitude and 41°35′~42°10′ north latitude, with a total area of 0.34×10^6 hm²[3,4]. Figure 1 shows an overview of the Taipusi Banner study area taken at 114°53′ east longitude and 41°40′ north latitude.

![Fig. 1 Overview of the study area](image)

In recent years, the ecological environment of Inner Mongolia has been in crisis, and many regions are facing threats from decreasing waters, increasing saline-alkali land, and raging sand cities. The situation in Taipusi Banner is not optimistic. Due to climate warming, soil erosion has increased, grass grazing has been overloaded, and grass has been degraded to varying degrees, causing widespread rodent damage. According to the information from the Grassland Research Institute, in 2015, Taipusi Banner grassland rodent damage occurred to a moderate degree, with an area of about 3×10^4 hm² [5].

2.2. Drone selection

Choose the DJI Mavic Mini drone that is light, easy to carry, simple to operate, and has better performance (Figure 2). The drone is equipped with a 1/2.3-inch CMOS image sensor, which can stably shoot 2.7K high-definition video and 12-megapixel photos. The remote controller works with the aircraft body to display high-definition images in real time on mobile devices through the DJI Fly App.

![Fig. 2 Royal Mavic Mini drone](image)

2.3. Data set production

The data used in the experiment are real grass mouse hole images SH taken by Mavic Mini drone and OPPO mobile phone with 1280*720 pixels, and processed by video frame extraction, manual screening and annotation. The production process of the SH data set is shown in Figure 3: First, use the labelImg tool to mark the mouse holes on the pictures, create an xml file for each picture, and save and record the position of all the marked mouse holes on the picture on the image. The coordinates are made into a data set in the standard VOC2007 format. Finally, the xml file is converted into a binary tfrecord format that occupies a small memory and has a fast-reading speed.
3. Mouse hole detection principle
In recent years, with the continuous development of computer vision technology, object detection research has set off another climax in the field of deep learning, which has been widely used in traffic information, face recognition and other aspects. The difference between target detection and image classification is that it not only has to determine what target is in the image, but also is responsible for accurately finding the location of the region of interest and using a rectangular frame to frame it[6].

According to whether it is necessary to extract candidate regions, the target detection based on deep learning can be divided into two categories: two-stage detection based on region selection and single-stage detection based on regression. The most commonly used algorithms in two-stage convolutional neural networks are: R-CNN, Fast R-CNN and Faster R-CNN. The single-stage target detection algorithm completely eliminates the idea of candidate regions and RPN, and directly performs regression and classification in a network, which greatly reduces the repetitive operation in the detection process, so the detection speed has been significantly improved, mainly represented by YOLO, SSD [7].

3.1. Analysis of Faster R-CNN Rat Hole Target Detection Principle
The R-CNN algorithm first uses a selective search method to extract multiple regional candidate frames in the image, and then normalizes and zooms each candidate frame to unify the size of the candidate area, and then performs CNN network extraction features and SVM classification processing. In 2015, Girshick[8] and others proposed Fast R-CNN, which solved the problem of different sizes of feature maps generated by inputting images of different sizes by increasing regional pooling, and avoided the loss caused by R-CNN zooming pictures[9]. However, Fast R-CNN's selective search method to generate candidate frames greatly hinders the improvement of accuracy. In response to this problem, in the same year, Ren [10] et al. proposed the Faster R-CNN algorithm, which no longer uses the selective search method in Fast R-CNN, and added RPN (Region Proposal Network) to generate candidate boxes, which not only reduces the time to generate candidate frames also improves the detection performance of the model. The target detection framework of Faster R-CNN is shown in Figure 4.

Although the two-stage target detection algorithm has made great progress in the detection task, the real-time performance is still not up to the requirement due to the complexity of the model.
3.2. The principal analysis of SSD mouse hole target detection

- SSD target detection process
  The SSD algorithm is a one-stage algorithm, and its English name is called Single Shot MultiBox Detector. The SSD algorithm uses a multi-scale detection method, which does not use a fully connected layer for prediction, but directly uses a convolutional neural network to detect pictures [9]. The SSD target detection process is shown in Figure 5:

![SSD target detection process](image)

**Fig. 5 SSD target detection process**

- SSD network structure
  As shown in Figure 6, the SSD network structure is mainly composed of two parts: The first part is to replace the fully connected layers FC6 and FC7 with the VGG-16 basic network of the convolutional layers Conv6 and Conv7 to extract the preliminary features of the input image; The second part is a set of pyramid-structured convolutional networks, used for multi-scale feature detection, and feature extraction of the feature layer generated by the first part of the network [11,12].

![SSD network structure](image)

**Fig. 6 SSD network structure**

3.3. Analysis of Improved SSD Rat Hole Target Detection Scheme
The SSD algorithm selects a series of candidate frames of different sizes on the six-layer feature maps of different scales to perform multi-scale prediction. Although multi-scale feature map prediction can achieve an effect similar to the image feature pyramid structure, its shallow feature map may lack the semantic information of the target because the receptive field is too small, causing the SSD algorithm to detect small targets obviously inferior to large targets. The target detection effect is good, and the detection accuracy of the algorithm still has room for improvement. At present, most of the algorithm improvements that help improve the detection accuracy of small targets are realized at the expense of the overall detection accuracy. Therefore, if we want to ensure the overall detection accuracy and improve the detection ability of small targets, we should start with how to improve the feature extraction of the model for small targets and increase the feature information [13,14].

The residual unit structure is shown in Figure 7:
Its idea is relatively simple, that is, on the standard antecedent convolutional network, a jump is added to bypass the connection of certain layers. Each jump is the output of the convolution block plus the input of the previous convolution block. The use of residual structure greatly reduces the difficulty of training deeper neural networks, and also significantly improves the accuracy.

4. Experimental results and analysis

4.1. Data set design
The 5181 grass mousehole data samples are divided into training set, test set and validation set according to the ratio of 8:1:1.

4.2. Experimental environment construction
The hardware and software environment used in the experiment is shown in Table 1.

| Category                  | Environmental conditions          |
|---------------------------|----------------------------------|
| CPU                       | Intel(R) Core (TM) i7-8700K      |
| Graphics card             | Nvidia GTX 1080Ti                |
| RAM                       | 64GB                             |
| Main frequency            | 3.70 GHZ                         |
| Hard drive capacity       | 1TB                              |
| Operating system          | Linux /Windows                   |
| Deep learning framework   | TensorFlow /PyTorch              |
| Programming language      | python                           |

4.3. Evaluation index
Similar to the machine learning problem, target detection uses precision rate, recall rate, etc. as its evaluation indicators. In the process of detecting mouse hole targets in the grass, we have four different outputs: the correct detection of the mouse hole target is called TP (True positives), the wrong detection of the mouse hole target is called FP (False positives), and the correct detection is called FP (False positives). Non-rathole targets are called TN (True negatives), and the wrongly detected non-rathole targets are called FN (False negatives), as shown in Table 2.
Tab. 2 The Evaluation index

| Predicted class | Actual class | + (Positive example) | - (Negative example) |
|-----------------|--------------|----------------------|----------------------|
| + (Positive example) | TP           | FN                   |
| - (Negative example) | FP           | TN                   |

With the four detected conditions, it can be explained as follows: TP is the number of targets whose detection results match the label in target detection; FP is the number of targets that are falsely detected; FN is the number of targets that are missed; TN is a true negative example. It is meaningless and generally not used.

Based on the judgment of true positive cases, false positive cases and false negative cases, we can get the precision (Precision), recall rate (Recall) and F value, among them:

Precision represents the ratio of the number of correctly detected targets to the total number of detected targets, which can show how the algorithm network model distinguishes the targets, as shown in formula (1).

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$  \hspace{1cm} (1)

The recall rate (Recall) represents the ratio of the number of correctly detected targets to the number of real frames (manually labeled), and expresses the query situation of the algorithm network model for the target, as shown in formula (2).

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$  \hspace{1cm} (2)

F value is a weighted average of model accuracy and recall rate, emphasizing the importance of smaller values, see formula (3). The maximum value of the F value is 1, and the minimum value is 0. Among them, 1 means that the output of the model is good, and 0 means that the output of the model is poor.

$$F = \frac{2PR}{P + R} \times 100\%$$  \hspace{1cm} (3)

Precision measures the accuracy of the detection results, and high Recall represents a lower missed detection rate.

4.4. Performance comparison

The 518 test set image data sets were tested on the trained Faster R-CNN and the improved SSD target detection model. The test results are shown in Table 3.

Tab. 3 Test result statistics

| Number of rat holes | TP  | FP  | FN  |
|---------------------|-----|-----|-----|
| Faster RCNN         | 505 | 8   | 13  |
| Improve SSD         | 500 | 6   | 18  |
| Artificial          | 516 | 2   | 2   |
According to the test results in Table 3, the accuracy rate, recall rate and F value of various detection methods are calculated respectively, as shown in Table 4.

| Model          | Ratio(%) | P    | R    | F    |
|----------------|----------|------|------|------|
| Faster R-CNN   | 98.4%    | 97.5%| 97.9%|
| Improve SSD    | 98.8%    | 96.5%| 97.6%|
| Artificial     | 99.6%    | 99.6%| 99.6%|

It can be seen from Table 3 and Table 4 that the detection accuracy, recall rate and F value of Faster R-CNN are similar to those of the improved SSD, but the time taken to detect each picture is 5 times (approximately) that of the improved SSD. As shown in Figure 8. Although manual detection can make the accuracy rate, recall rate and F-value all reach 99.6%, the manual detection time will be longer, and the detection time for a picture is about 1-2 seconds. In summary, the improved SSD target detection effect is better than other detection methods.

![Performance comparison histogram](image)

**Fig. 8 Performance comparison histogram**

4.5. **Model measurement**

From the grass mouse hole detection effects of the above different target detection algorithms, it can be found that the improved SSD target detection model not only ensures the accuracy of the mouse hole target detection in the grass mouse hole image, but also improves the detection speed of each image. Research purposes. Next, 5 pictures with mouse holes and 1 picture without mouse holes will be randomly selected from the grass pictures that are not involved in the construction of the network model to test the improved SSD target detection model. The test results are shown in Table 5.
From the test results in Table 5, it can be seen that the improved SSD target detection model can basically accurately realize the detection of grass mouse hole targets.
5. In conclusion
The combination of machine vision and unmanned aerial vehicle used in Inner Mongolia grassland rat hole detection can improve detection efficiency on the basis of ensuring accuracy and make up for the long detection cycle of traditional manual survey methods. Compared with similar studies, the novelty and innovation of this article are mainly reflected in the high time-efficiency performance. At the same time, the wrong detection and missed detection of grass mouseholes can still be further improved or improved by improving the non-maximum suppression algorithm.

Acknowledgments
Natural Science Foundation of Inner Mongolia Autonomous Region (2019MS06023); Science and Technology Research Project of Inner Mongolia Autonomous Region (2019GG334, 2019GG337)

References
[1] Zhou Xiaolin, An Ru, Chen Yuehong, Ai Zetian, Huang Lijun. Research on remote sensing recognition of ratholes in the typical area of the Three Rivers source [J]. Journal of Subtropical Resources and Environment, 2018, 13(04): 85-92
[2] Sun Di, Ni Yifei, Chen Jijun, Abdulwari, Zheng Jianghua. The application of unmanned aerial vehicle (UAV) low-altitude image monitoring of yellow rabbit-tailed rat mouse hole [J]. China Plant Protection Guide, 2019, 39(04) :35-43
[3] Bao Xuefeng. The response of climatic productivity of natural grassland pasture in Xilin Gol League to climate change [D]. Inner Mongolia Agricultural University, 2015
[4] Miao Lijuan, Xu Xia, Lan Yufang, Cui Xuefeng. Evaluation of the impact of the project of returning grazing to forest (grass) on the human environment in the agro-pastoral ecotone in northern China: Taking Taipusi Banner as an example [J]. Journal of Beijing Normal University (Natural Sciences) Edition), 2013, 49(01): 46-50
[5] Wang Bifei, Zou Bo, Niigao Shen, Wang Qing, Zhou Enfang, Li Weiwei, Cao Qiufen, Chen Jie, Zhang Xiaofang, Yang Mingfeng. Grassland rodent pests and their prevention and control [J]. Agricultural Technology and Equipment, 2015(11): 43 -47
[6] Wu Xue, Song Xiaoru, Gao Song, Chen Chaobo. Overview of target detection algorithms based on deep learning [J]. Sensors and Microsystems, 2021, 40(02): 4-7+18
[7] Li Niuniu. Research on traffic sign recognition method based on convolutional neural network [D]. Xi'an University of Science and Technology, 2020
[8] Girshick Ross. Fast R-CNN [J]. Computer Science, 2015
[9] Xu Rong. Research on small target detection algorithm based on YOLOv3 [D]. Nanjing University of Posts and Telecommunications, 2020
[10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149
[11] Wang Ziyu. Research on tea bud detection technology based on image [D]. Shenyang University of Technology, 2020
[12] Liu Pan. Research on a pedestrian detection method to improve SSD [D]. Xi'an University of Technology, 2020
[13] Li Maopeng. Research on target detection algorithm and pruning optimization based on improved SSD [D]. Nanjing University of Posts and Telecommunications, 2020
[14] Zhang Jiaming. Research on traffic sign detection and recognition based on deep learning [D]. Dalian Maritime University, 2020