AUTOMATIC VIDEO DETECTION FOR GOLDEN MONKEYS IN SHENNONGJIA NATIONAL PARK, CHINA

Haigang Sui [1], Tianyi Wei [1], Jindi Wang [1] *, Li Hua [2], Na Xiong [1]

[1] State Key Laboratory of Information Engineering in Surveying, Wuhan University, Wuhan, China - wangjd@whu.edu.cn
[2] College of Resources and Environment, Huazhong Agricultural University, Wuhan, China - huali@mail.hzau.edu.cn

KEY WORDS: Golden Monkeys, Infrared Camera-trapping, Video Detection, Computer Vision, Deep Learning.

ABSTRACT:

Infrared-triggered cameras serve as essential tools for wildlife resource surveys, allowing camera traps to capture wildlife activity while minimizing the impact on the ecosystem. Traditional monitoring methods heavily rely on human resources for visual discrimination, which is inefficient and susceptible to environmental influences. Therefore, this paper focuses on Shennongjia National Park as a case study to address the limitations of traditional approaches. We constructed a Shennongjia wildlife object detection dataset using video data from the fixed-erected infrared cameras and proposed a supervised learning-based automatic monitoring method for golden monkeys, aiming to achieve intelligent target detection. With these processing results, we were able to capture golden monkeys' tracks and conduct statistical analysis of their life range and distribution characteristics. Through the integration of sensors and deep learning techniques, we developed a golden monkeys detection monitoring system to visualize the monitoring results and assess the spatial and temporal distribution characteristics of golden monkeys' activities.

1. INTRODUCTION

Golden monkeys play a crucial role in ecosystem research, and accurate monitoring of primates is essential for wildlife research, conservation and management decisions. In 2016, during the 40th General Assembly of the UNESCO World Heritage Committee, Shennongjia, Hubei Province, was officially inscribed on the World Heritage List and was honoured as a "World Natural Heritage Site". The Shennongjia golden monkey is an indispensable species in the ecological succession and material cycle of the Shennongjia ecosystem. With the increasing level of industrialization in cities and towns at this stage, the destruction of natural resources becomes inevitable, highlighting the remarkable impact of Shennongjia's ecosystem on the global ecosphere. How to protect the golden monkeys and other wildlife in Shennongjia in a better way has also become an essential issue to which the local government is paying more and more attention.

The early recognition and tracking of animals were mainly achieved through the wireless radio frequency identification technology (RFID). However, for wild animals, electronic tags may have an incalculable impact on their return to tribes, hunting, and escape from their natural enemies. Therefore, non-invasive infrared trigger cameras have become one of the main approaches of wildlife resources investigation. Camera traps are remotely operated cameras that use motion sensors, infrared detectors, or other light beams as triggering mechanisms, and are often used to photograph animals that live in the wild out of the way (Villa et al. 2017; Giraldo-Zuluaga et al. 2017). In recent years, Shennongjia National Park has been strengthening the construction of monitoring systems, establishing a relatively complete infrastructure to achieve a 24/7 and multi-scene wildlife monitoring. As shown in Figure 1, infrared cameras can capture the activity images of different wild animals in Shennongjia during different seasons, day and night.

Figure 1. Images captured by infrared cameras

Even though, the existing monitoring system lacks sufficient intelligent, heavily relying on human effort to distinguish animals from videos. This approach not only wastes manpower but also introduces significant delays and is highly influenced by environmental conditions. Therefore, it is urgent to develop intelligent monitoring techniques and efficient data processing methods for golden monkeys. Computer vision, a rapidly developing field for recent years, offers a powerful tool for processing massive amounts of video image data. The development of image object detection can be divided into two stages, the early stage is dominated by traditional image processing and machine learning algorithms, which are extremely dependent on the design of manual features, and the scope of application of algorithms was limited (Tabak et al. 2019). However, over the past decade, with the exponential growth of computer computing power, neural networks have once again returned to this filed (Rawat et al. 2017), and deep learning breed many object detection models, of which there is no lack of accuracy and efficiency we focused on (Krizhevsky et al. 2012; Chen et al. 2014).

Wildlife including golden monkeys are mostly active in mountainous areas, which add considerable difficulties to their

* Corresponding author

detections. Their activities are random, and infrared cameras can rarely capture frontal shots images. Meanwhile, small distant animals are difficult to detect due to environmental influences. Therefore, the unimproved universal object detection model tends to give a false detection and omissions. Based on above, this paper proposes an automatic detection method for golden monkeys in Shennongjia National Park, employing the single-stage object detection model YOLOv5 for intelligent extraction. To validate our method, we built a new dataset - Shennongjia Wildlife Images Dataset - for wildlife detection in Shennongjia. In order to detect a specific animal (e.g. golden monkeys) from massive pictures taken by infrared camera, we developed a system that integrates deep learning and sensor technologies for monkey detection and monitoring. By combining sensor locations for fixed-point monitoring with target detection from videos, we conducted statistical analysis and visualization of the golden monkeys' quantity. This allowed us to derive the spatial and temporal distribution characteristics of golden monkeys.

2. MATERIALS AND METHODS

To address the problems of intelligent monitoring, we proposed a method shown in Figure 2, consisting of three parts. Firstly, we extract frames from infrared cameras videos and annotate them both manually and automatically to build a dataset; Then, we train the dataset repeatedly through the improved model, and get a most suitable detection model for the golden monkeys in Shennongjia area; By integrating the deep learning model and sensors, we can verify and test the images taken by infrared camera, visualize the results, and study the quantity and distribution characteristics of golden monkeys by maps, charts and other forms.

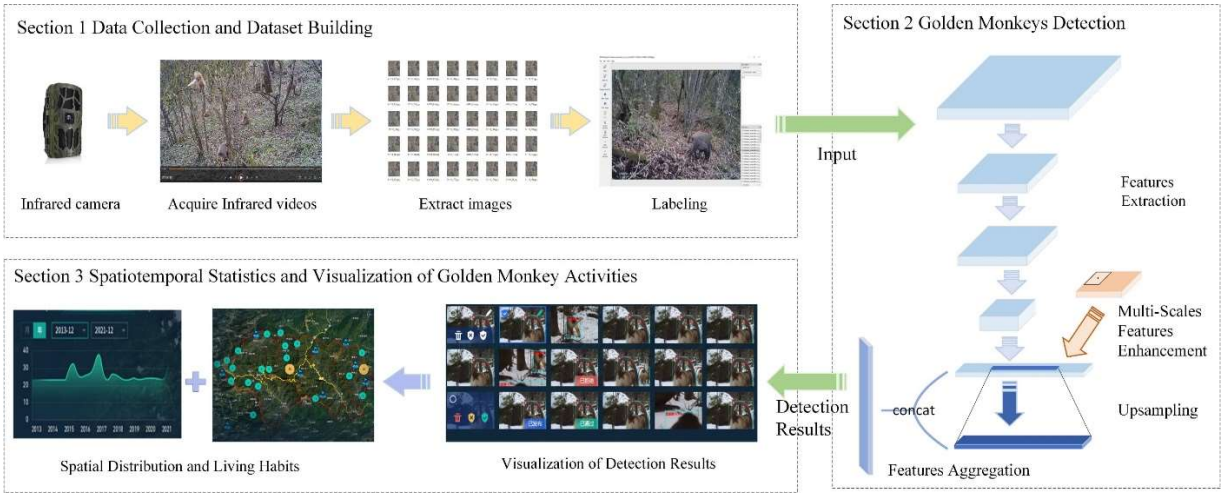

Figure 2. Overview of golden monkeys detection method

2.1 Material

2.1.1 Study area

The primary area of the Shennongjia National Park is situated within the Shennongjia National Nature Reserve (Shennongjia Forest Area) in Hubei Province, also includes part of the Badong Golden Monkeys National Nature Reserve (Badong County). It is located on the eastern edge of the second terrain in China, and extends in an east-west direction, with a high southwest and low northeast terrain and deep canyons. The highest peak in Hubei Province, Shennong Peak, stands at an elevation of 3,106.2 meters above sea level, making it the highest point in Central China. The complex topography, huge height difference, various soil types and climatic conditions have given birth to multiple habitat types and rich biodiversity, and its ecosystem is complete and the largest area of primary forest distribution in Central China.

Figure 3. Topographic Map of Shennongjia

2.1.2 Data collection

Image annotation plays a crucial role in computer vision. The goal of image annotation is to provide task-relevant and task-specific labels. This may include text-based labels (classes), labels drawn on images (borders), or even pixel-level labels. During the experiments, this project relies on image video data from Shennongjia for further processing and training. We use labelImg (an image annotation tool) for labelling. For the video data, we used inter-frame selection to convert it into image data and set different labels according to different animals to lay the data foundation for the subsequent target detection training. The labelling adheres to uniform rules, such as close to the edge, separating animals that are independent but close together, and labelling all small targets that can be distinguished by the human eye as much as possible. Due to the large size of the dataset, manual labelling is very slow and time-consuming, so we put a portion of the labelled data into the model, and the results were checked and supplemented to form the complete dataset. After correcting the detected error categories and error frames, and improving the original dataset, the final dataset was completed with 16 animal categories, 89,406 images and 107,986 instances (one images may has several animals), as shown in Figure 3 and table 1. It should be noted that the data labels formed into the dataset are not arranged in order. Therefore, although the maximum number of data labels is 20, the dataset has only 16 classes.

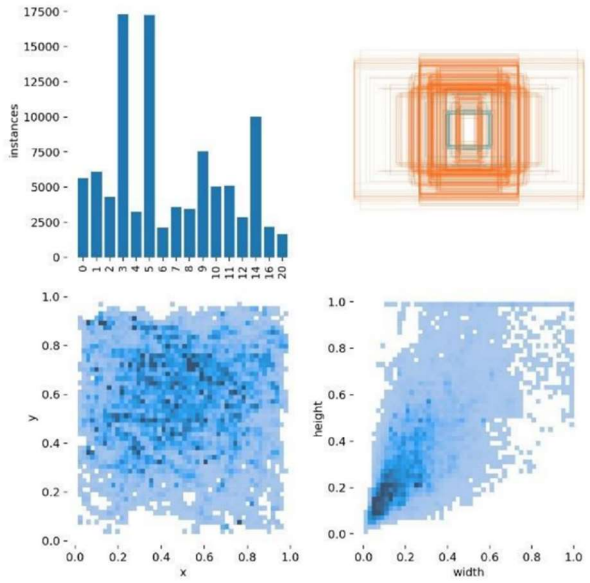

Figure 4. Dataset categories and target boxes

Labels	Classes	Instances
0	Golden Monkey	6239
1	Chinese Muntjac	6768
2	Tufted deer	4772
3	Sika deer	19200
4	Porcupine	3579
5	Wild Boar	19128
6	Hog Badger	2359
7	Black Bear	3988
8	Goral	3826
9	Yellow-throated Marten	8357
10	Golden Pheasant	5608
11	Tragopan Temminckii	5684
12	Rock Squirrel	3164
14	Macaque	11070
16	Mouse	2413
20	Ocelot	1831

Table 1. Dataset information

2.2 Methods

In this section, we will introduce a wildlife target detection method based on YOLOv5 (Jocher et al. 2020), which adds new and smaller detection scales to the original three detection scales to detect animals that are far away from the camera or smaller in size. This model should not only realize species classification in many videos of Shennongjia National Park, but also be able to find golden monkeys from images. YOLOv5, developed and maintained by Ultralytics, is based on a PyTorch implementation that makes it easier to train and deploy. Currently YOLOv5 is available in a variety of models of different sizes, with the advantages of speed, accuracy, and small size. The original network has three layers of scale features to, which make it possible to detect most of animals. However, there are still some animals that cannot be detected by the YOLOv5 network due to their distance from infrared cameras or their small size. Thus, we add a new scale to the original structure for detecting smaller-scale animals, and obtains a new feature map through expansion, concatenation, and fusion, as Figure 5 shown.

Please note that some modules have been omitted in Figure 5 for a clearer presentation. We will focus on explaining the modified modules, while leaving the unchanged YOLOv5 modules without further elaboration. In the process of extracting small-scale feature layers, the input images will be first convolved 4 times with C3 module, and then continuously convolved and upsampled after passing through SPPF module, fused with the previously obtained feature layer. On this basis, in order to obtain new feature maps that can detect smaller animals, we extended a new scale of 160×160 on the existing detection layer and incorporated an additional upsampling operation as part of the fusion process. By concatenating and fusing the obtained upsampling features and 160 ×160 feature layers, we can finally get a new scale feature map. It should also be noted that in the three detection scales of the original YOLOv5, the fused feature layers of large and medium scales will remain unchanged, while the fused feature layers of the small scales, as they are no longer the minimum scale, need to obtain convolutional layers from the new smaller scales we have added. In addition, we also added three additional anchor boxes for detecting small animals by clustering our own dataset. The improved model has increased computational complexity and detection time, but the accuracy of smaller targets has been improved greatly.

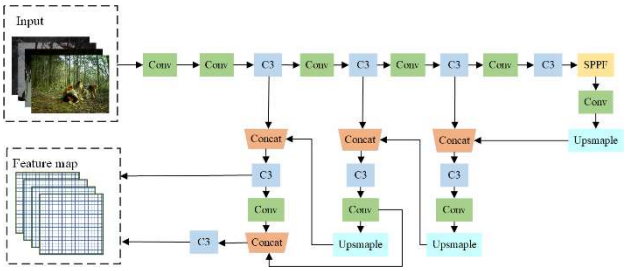

Figure 5. Smaller-scale feature maps extraction

3. EXPERIMENTS AND APPLICATION

3.1 Experiment

In this paper, several evaluation metrics are employed, including precision (P), recall (R), mean average precision (mAP), and frames per second (FPS). TP represents the true positive value (number of positive samples correctly identified), FP represents the false positive value (number of negative samples for false positives), TN represents the true negative value (number of negative samples correctly identified), and FN represents the false negative value (number of missed positive samples). The accuracy and recall are defined as shown in equation (1)(2):

$$Precison = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

The curve plotted between the two is the P-R curve. the area enclosed by the P-R curve is AP, as shown in equation (3). mAP@.5 refers to the mAP when IoU=0.5. mAP@0.5:0.95 refers to the average mAP at different IoU thresholds (from 0.5 to 0.95 in steps of 0.05), which can be calculated using equation (4).

$$AP = \int_0^1 P(R) \quad (3)$$

$$mAP = \frac{\sum AP}{N(class)} \quad (4)$$

Based on the dataset collection process in 2.2, the first batch of about 40,000 initial datasets is first trained for 300 iters. It can be seen that the initial dataset is highly susceptible to overfitting due to the uneven distribution of each category. As the equation defined, the closer the P-R curve is to the coordinate (1,1), the better the algorithm performance is indicated. As can be seen in Figure 6, the training effect of the uncorrected data is not satisfactory, which is also confirmed by the confusion matrix in Figure 7. As for the test results, due to imbalance of samples, on one hand, some animals are incorrectly detected as wild boars with much larger number of data set than other categories; on the other hand, several kinds of animals are often missed when they coexist in the same image as other species.

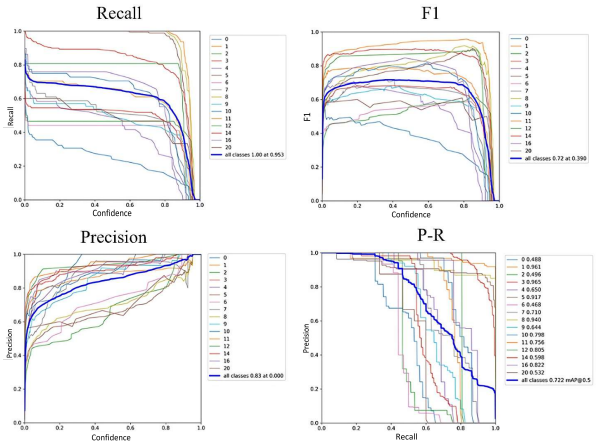

Figure 6. Training results after the first experiment

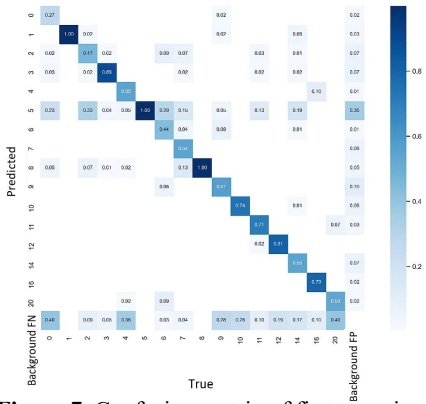

Figure 7. Confusion matrix of first experiment

As the uncorrected dataset trained initially falls short of the desired performance, it becomes evident that further processing is required before incorporating the dataset into training. Therefore, we train the remaining unlabelled dataset by using this network, and the results obtained from it are checked and modified manually, and then the detected error categories and error boxes are corrected separately to finally obtain the complete dataset of about 90,000 images. These 90,000 images were put into training again.

The second training was chosen with fewer training epochs and the improved yolov5 network to obtain a better detection result for smaller animals. The modified dataset somewhat balances the great differences between the number of various types of labels and avoids the serious overfitting phenomenon. The evaluation indicators in Figure 8 and 9 also show that the modified dataset substantially improves the model accuracy. Table 2. shows the values of indicators in twice trainings. The improved accuracy of the training is reflected in the results, and the detection of animal targets in the images is more accurate, and the number of wrong and missed detections is greatly reduced.

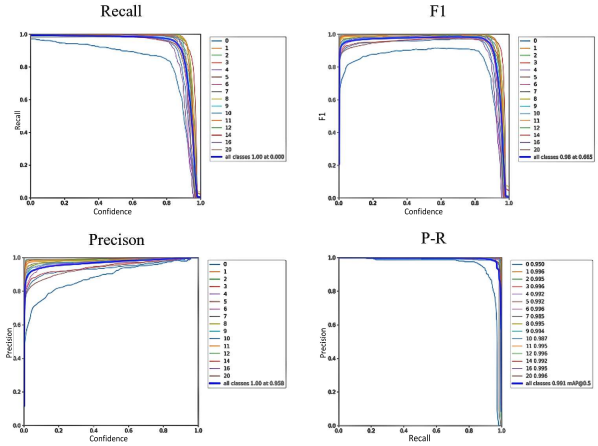

Figure 8. Training results after the second experiment

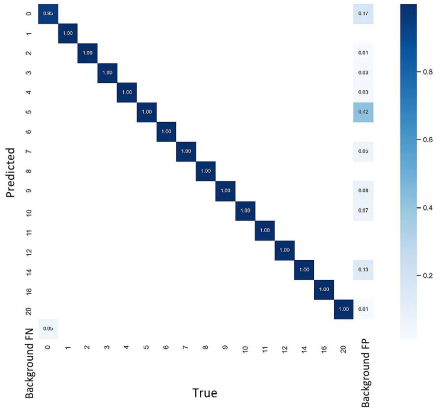

Figure 9. Confusion matrix of second experiment

Exp.	P	R	mAP@.5	mAP@.5:.95	FPS
1	0.832	0.671	0.722	0.632	32.32
2	0.985	0.983	0.991	0.897	18.08

Table 2. Comparison of accuracy between two training sessions

Besides, we trained the modified network for 200 rounds same as the first experiment to evaluate the effect of models, and obtained the training loss curves shown in Figure 10. The blue curve is the first training, and the red curve is the modified second training, from left to right are the border regression loss, classification probability loss and confidence loss respectively. The initial loss of two training is relatively similar, but as the batch increases, the loss of modified algorithm is smaller than the former.

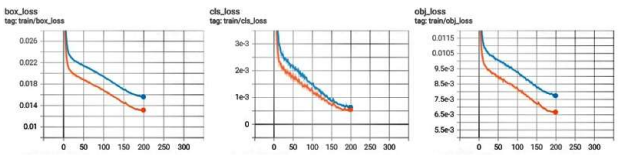

Figure 10. Training loss comparison curve

In order to show the detection effect, golden monkey videos in the test set are randomly selected for testing. The effect shown in Figure 11 can present that although some animals cannot be detected perfectly due to the large changes in posture, the distant

animals were detected more accurate in the images after the modified model training.

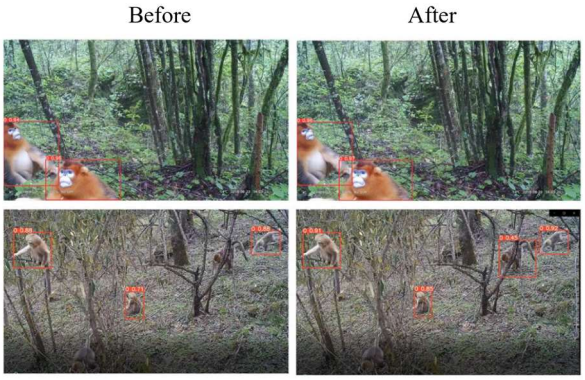

Figure 11. Detection results comparison

3.2 Application

We develop a management and intelligent analysis system based on a B/S architecture, combining with the deep learning model and infrared cameras.

The system function consists of the following parts:
(1) Data management. This part is responsible for acquiring, inputting and outputting the raw video data. And it also responsible for cleans and handle images from the video.
(2) Monkey detection analysis. This part is the core of the system and is responsible for detecting golden monkeys.
(3) Visualization interface. This module is in charge of visualizing the results, intuitively outputting analysis results, including data chart statistics and map display, to provide support for decision-making.

3.2.1 Data management: Data management includes camera information, animal category classification, task assignment, training samples, model management and so on, mainly intended to realize the entry management of animal-related information in Shennongjia National Park and the storage of animal extraction training and related information. As the infrared camera video of wild animals is still essentially fixed-point monitoring, the camera number can help identify the location and range of wild animal activities. Through animal category classification, staff can effectively identify and manage wildlife species in the Shennongjia area. At the same time, through the management of training samples and models, the timely update of models can be realized, which helps detect more accurately and more targeted to the Shennongjia area.

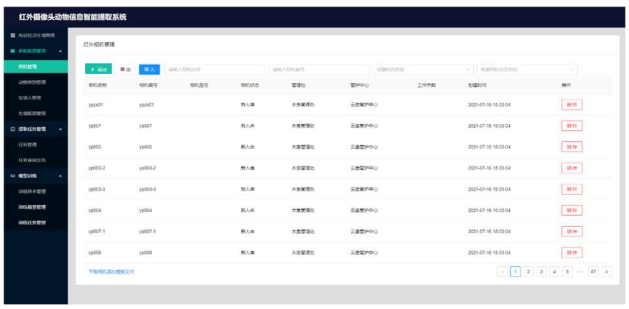

Figure 12. Data management interface

3.2.2 Monkey detection analysis: This function includes visual uploading detection videos, extracting tasks, and reviewing extraction results, etc., which provide data results for the final distribution analysis. The resource list on the left side can view the sources, deployment points and shooting dates of animal detection videos. The bottom results show the results of automatic model detection, while the system provides a manual correction function in the meantime, offering administrators the option to check the accuracy of automatic monitoring. The correct results will be retained to provide data samples for subsequent improvements, and ambiguous or incorrect samples will be further processed or deleted. The list on the right is the animal activity statistics, showing the type and number of wild animals active at a particular date and time.

Figure 13. Interface of monkey detection

3.2.3 Visualization: This function will obtain the distribution of animals in terms of category, quantity and space based on the extracted results and spatial and temporal information. The charts on the left side will provide statistics on some basic conditions, such as the number of video files, animal types, camera deployment plan and erection, etc. These statistics can help understand more intuitively the scale of the infrastructure for wildlife monitoring in the Shennongjia area. On the right side of the system is the statistical analysis of the monitoring results. Through these charts, we can have a clearer understanding of the number and types of species in this area in the past.

Figure 14. Visualization chart interface

The main part of the system, the map of monkey monitoring in Shennongjia, shows the range and activity of golden monkeys in a more intuitive way. At the same time, by clicking on the map to see the specific time of golden monkeys, we can have a general understanding of the activity's habits of golden monkeys in the area, and thus make more scientific protection.

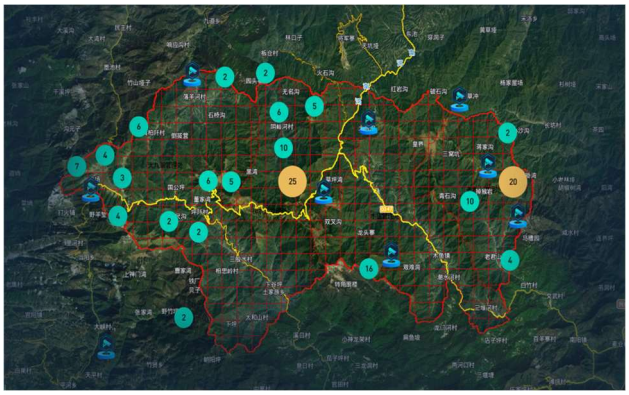

Figure 15. Distribution of golden monkeys in shennongjia national park

4. CONCLUSION AND FUTURE WORK

Thanks to the continuous development of sensor technology and computer technology, we can more easily collect a large number of animal movement data. In this study, our objective is to achieve intelligent monitoring of golden monkeys in the Shennongjia National Park, enabling us to uncove to uncover the activity patterns of local wildlife. This research holds significant importance for biologists studying the wildlife in this region.

For this purpose, we need to be able to intelligently and accurately extract and detect golden monkeys from videos of so much wildlife. Considering the variability of animal size and the limitation of infrared camera monitoring, we improved Yolov5 to make it more suitable for animal detection in Shennongjia Nation Park. At the same time, we integrated sensor technology with deep learning to develop a golden monkeys management and intelligent analysis system to integrate infrared video and process large amounts of data efficiently. Our system assists monitoring operators in obtaining information on golden monkey species, numbers and activities to study their habits and timely rescue operations. However, wildlife conservation should not be limited to species identification alone . To gain a comprehensive understanding of wildlife ecology and habitats, it is necessary to further achieve the tracking of specific target individuals. It is our hope that as the model continues to be improved and refined, the research will continue to offer better solutions for transitioning from analyzing animal groups to tracking individual animals.

REFERENCES

Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. 2014. Deep convolutional neural network based species recognition for wild animal monitoring. In 2014 IEEE international conference on image processing (ICIP), pp. 858-862

Giraldo-Zuluaga, J. H., Salazar, A., Gomez, A., & Diaz-Pulido, A. 2017. Recognition of mammal genera on camera-trap images using multi-layer robust principal component analysis and mixture neural networks. In 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 53-60.

Jocher, G, 2020. YOLOv5 by Ultralytics. Code repository https://github. com/ultralytics/yolov5.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Rawat, W., & Wang, Z. 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9), 2352-2449.

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., VerCauteren, K. C., Snow, N. P., & Miller, R. S. 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10(4), 585-590.

Villa, A. G., Salazar, A., & Vargas, F. 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological informatics*, 41, 24-32.