Optimized Detection Method for Siberian Crane (Grus leucogeranus) Based on Yolov5

Linlong Wang, Huaiqing Zhang, Tingdong Yang, Jing Zhang, Zeyu Cui, Nianfu Zhu, Yang Liu and Yuanqing Zuo

April 8, 2021

Optimized Detection Method for Siberian crane (*Grus leucogeranus*) Based on *Yolov5*

Linlong Wang
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
wangll1988@caf.ac.cn

Huaiqing Zhang
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
zhang@ifrit.ac.cn

Tingdong Yang
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
yangtd@ifrit.ac.cn

Jing Zhang
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
zhangjing@ifrit.ac.cn

Zeyu Cui
Key Laboratory of Forest Management and Growth Modelling, NFGA, Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
645011764@qq.com

Nianfu Zhu
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
m13477032471@163.com

Yang Liu
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
liuyangifrit@gmail.com

Yuanqing Zuo
Key Laboratory of Forest Management and Growth Modelling, NFGA ,Institute of Forest Resource Information Techniques
Chinese Academy of Forestry
Beijing, China
13705622833@163.com

***Abstract*—In our study, we have analyzed the detection accuracy of the *YOLOv5* model based on different training datasets and iteration times, to select the optimal dataset for the following research and build a *YOLOv5*-based optimization detection model of Siberian crane by using deep learning method and training datasets collected by network crawling, panoramic cameras and *SLR* cameras. The results show that, (i) the *mAP* values of the six training datasets are Train_4 > Train_2 > Train_1 > Train_3 in turn. The *mAP* value of Train_4 was biggest, reaching to 90.9%, which was much higher than other training dataset, and indicate that the model detection accuracy of the training dataset mixed with ordinary field photos and network photos was much higher than that of the training dataset from a single source image; (ii) when the iteration times reach to 40000times, *YOLOv5* model can completely converge, and *mAP* value reached to 81.3%, total loss value 0.0357; (iii) According to the result of the existing model test, we found that the model can also have an effectively performance of detection in the complex environments, such as pictures exist the problems of multi-objective small objects and occlusions, similarity in color between the object and background, and Siberian cranes are in different activities such as flying, falling, foraging, playing etc.**

Keywords—Siberian crane; deep learning; object detection; YOLOv5

I. INTRODUCTION

The Siberian crane (*Leucogeranus leucogeranus*), a large wading bird, has been identified as one of the critically endangered species on the International Union for Conservation of Nature (IUCN) 2018 Red List, with a total number of about 3500-4000 worldwide [1]. They have a long history, which is known as the "living fossil" of birds, and have important value of historic and culture. However, due to habitat destruction, human hunting, environmental degradation and other problems, the number of Siberian crane has been declining in recent years, even under the threat of extinction, which has brought adverse effects on biological species diversity and genetic diversity, and destroyed the ecological balance of nature [2,3]. Therefore, it is urgent to take scientific and effective measures to protect the survival of Siberian crane. Monitoring wild animals to identify their population and habits is considered to be important for the conservation and management of ecosystems.

Currently, the methods of dynamic monitoring of wildlife population include mainly intuitive statistical method, such as the total ground count method[4], the line-transect count method[4], the dropping count method[5], and aerial count method[6,7], and these methods are based on human observation to intuitively count the animals in local areas, and then use these information to estimate the size of the population in the whole area[8]; computer image processing methods, such as automatic counting based aerial count method, and the method uses aerial photography and UAV to obtain image dataset, and then uses image processing algorithm and computer system to process a large number of images quickly to realize automatic counting function, which has the characteristics of low time and price consumption and fast speed[9-13]. However, related studies only focus on limited number of images of specific species in a specific environment, and the number of images used is relatively small, so there are some limitations on the study of birds in different environments[14].

Object detection based on deep learning has become one of research hotspots in the field of animal monitoring in recent years. Compared with the traditional manual feature-based machine learning method, the deep learning method can learn features by itself from a given large amount of image data and has a good detection performance[15]. Among deep learning methods, convolutional neural network *CNN* is the most commonly used object detection method based on deep learning, because of the special architecture of network[16]. *CNN*-based object detection methods such as Region-based Convolution Neural Network (*R-CNN*)[17], *Fast R-CNN*[18], *Faster R-CNN*[19,20] consist of two stages: object classification and bounding box location. One-stage object detection methods, such as You Only Look Once (*YOLO*)[21], Single Shot MultiBox Detector (*SSD*)[22], and Retinanet[23], complete the bounding box location and object classification simultaneously. One-stage object detection method have a faster speed performance comparing to the two-stage methods. However, the performance in term of computing speed and accuracy for these methods is different because of the type of CNN architecture employed, e.g. *Alexnet*[19], *Googlenet* (Inception) [24], *VGGNet*[25], *Squeezenet*[26], *Resnet*[27], or *Densenet*[28,29]. With the development of various algorithms, high precision detection based on real time is the goal of detection. Some studies have shown that , compared with *Faster R-CNN* model, *YOLO* algorithm is faster and the the rate of Background check error is lower, so it is more popular in some application fields. However, there are also some challenges in the development of *YOLO*, such as simple training dataset, lacking of the details of the image feature extraction, poor performance in small object detection, lower value of precision and recall rate. To address these problems, it is urgent to collect the multi-source dataset for training and improve the performance of the *YOLO* model by improving the architecture of the model.

Many interests and research efforts have been devoted to monitor Siberian crane in Poyang Lake for various aims such as population and habitat investigate, and ecosystem conservation. In this study, based on the deep learning model of *YOLOv5*, we use traditional *SLR* cameras Panoramic camera to make the training dataset of Siberian crane in different environment, to have a knowledge of the influence of different source of images (network general panorama) on detection performance of Siberian crane, so as to select the best dataset for subsequent scientific research; and to build optimal model for the detection of Siberian crane in a complex environment.

II. Materials and Methods

A. Study area

The study area (116°13 '48 "-116 °54' 24" E and 28°21 '36 "-29 °3' 24" N) is located in Poyang Lake, a seasonal throughput lake and the largest freshwater lake in China[30]. Because of its rich plant, fish and other species diversity, this lake is important for the protection of global biodiversity and is an important winter migration destination for various migratory birds. Nearly 98% of the world Siberian cranes population, 50% of the white-naped crane (*Grus vipio*) population, and 60% of the wild geese population spend their winters in the *Poyang* Lake wetland[31]. Besides these species, there are several protected species such as grey goose, pheasant, crested duck, bean goose, white spoonbill, little swan, kestrels, green-winged teal *etc*.

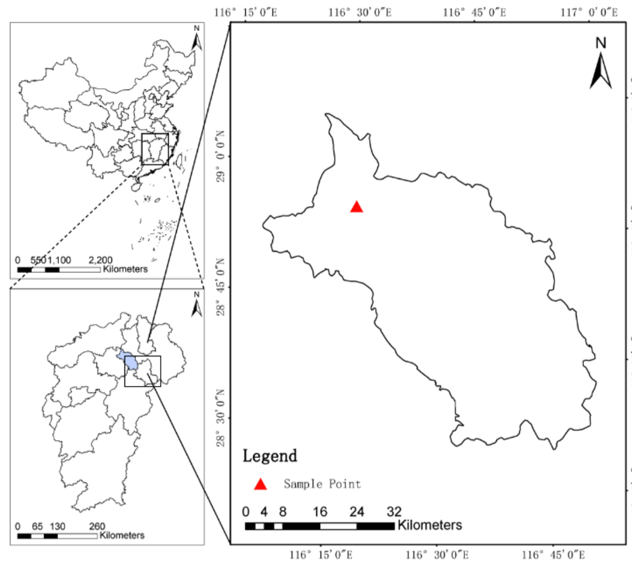

Fig. 1 The geographic location of site in Poyang Lake

B. Experimental data

The model of deep learning is a data-driven algorithm, and the performance of the model will be determined by the quality and quantity of datasets. Generally speaking, the richer the datasets are, the better the learning efficiency and the stronger the detection performance will be.

1. Data Collection and Processing

The collected datasets of Siberian cranes were derived from: (1) Internet datasets. We collected 1020 images of Siberian crane by using Python web crawler technology and finally only chose 200 images according to the quality of images; (2) Field ordinary datasets by *DSLR* cameras. We chose 200 pictures of Siberian crane by *DLSR* cameras, which of the resolution was 3984×2656. (3) Field panoramic datasets by *TITAN* which is panoramic camera from the company pf Insta360. We selected 200 pictures which of the resolution was 7980×3840. The selected images of Siberian cranes are characterized by uncertainty in number, posture and size, and difficulty in training with small object, unobvious features and cover-up in the cluster of Siberian cranes,. In order to address these problems, we spent a lot of time in choosing the proper images for datasets, which included pictures of single and group Siberian cranes, pictures of flying and standing cranes, pictures of foraging and playing cranes. The number of pictures with different categories was in a roughly balanced proportion.

The Siberian crane datasets were divided according to the proportion of training dataset 80% and test dataset 20%. In order to improve the number of training datasets of Siberian crane, we adopt data augment strategies which include *HSV* adjustment, random brightness adjustment, horizontal flip, rotation of specified Angle, Gaussian noise point, etc. which led to increase the detection accuracy, generalization and robustness ability of the model. After data enlargement, there are 3600 data pictures used for training, including 1200 network pictures, 1200 ordinary pictures and 1200 panoramic pictures from the field.

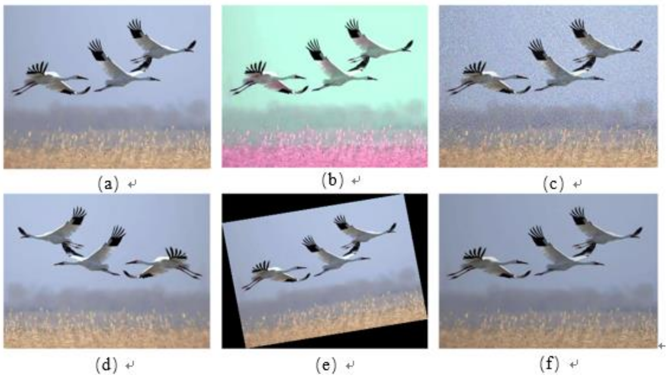

Fig. 2 Image preprocessing of Siberian crane (The letters of a-f means the different way of preprocessing of the Image. (a) represents the original image;(b) means the image under *HSV* regulation;(c) means the image under Gaussian noise;(d) means the image under horizontal inversion;(e) means the image under rotation by 10°;(f) means the image under random brightness adjustment.)

2. *The construction of different training datasets*

In order to address the problem of single defect of the existing animal training datasets and have a knowledge of the performance of object detection based on panoramic images, this paper will talk about the influence of the training datasets constructed from different sources of images on the detection performance of Siberian crane. According to the sources of photos, we classify the datasets are into six types (see table 1): (i) Train_1, means the training datasets which only consist of 1200 images from Internet under several ways of data augment. (ii) Train_2, means the training datasets which only contains 1200 ordinary images from the field; (iii) Train_3, means the training datasets which only consist of 1200 panoramic images from the field; (iv) Train_4, means the training datasets consisting of 1200 images from internet and 1200 ordinary images from the field; (v) Train_5, means the training datasets consisting of 1200 images from Internet and 1200 panoramic images from the field; (vi) Train_6, means the training datasets consisting of 1200 ordinary images and 1200 panoramic images from the field.

Table 1 The Number of images in the Training dataset

Index	Training Datasets	Sources of Datasets	The Way of Data Augment	Number of Images in Datasets
1	Train_1	Internet	HSV regulation, Gaussian noise, horizontal inversion, rotation, random brightness adjustment	1200
2	Train_2	Ordinary images		1200
3	Train_3	Panoramic images		1200
4	Train_4	Internet and Ordinary images		2400
5	Train_5	Internet and Panoramic images		2400
6	Train_6	Ordinary and Panoramic images		2400

C. *Research Approach*

1. *YOLOv5*

(1) Principles of *YOLOv5*

The model of *YOLOv5* extracts features from the input image through the feature extraction network, and obtains the feature map of a certain size. Then the input image is divided into *S*S* grid cells, and each grid is responsible for detecting the target falling into it, and finally the bounding box, positioning confidence and probability vector of all the targets contained in all grids are predicted once.

(2) Architecture and Components of the Network

The model of *Yolov5* was first released on GitHub in June 2020 by *Ultralytics*. The structure of the model is divided into two parts: (1) The Backbone part, including *Focus*, *BottleneckCSP*, *SPP* and other modules; (2) The Head part, consisting of *PANet* and *Detect* modules. The function of *Focus* module is used to reduce the calculation amount of the model to a certain extent in the process of image subsampling, and will not cause information loss, so as to retain complete image subsampling information for subsequent feature extraction.

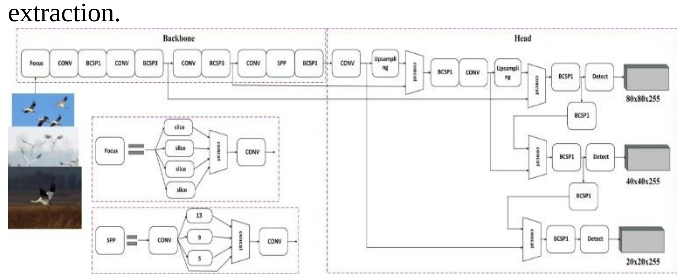

Fig.3 Architecture of *YOLOv5s*

The most important function of *SPP* module is to solve the problem that the size of input image is not uniform. In most networks of object detection, the fully-connection layer is generally used as the output layer in the end, which requires the size of the input image to be fixed. However, some image preprocessing operation, such as resize, crop and other operations, will cause a certain degree of image distortion, which will affect the final accuracy of detection. *SPP* module uses fixed block pooling operation, can achieve the same size of output for different size of input, so it can effectively avoid this problem. The fusion of different size features in *SPP* module is beneficial to the situation of that the image to be detected were in a big difference, especially for complex multi-object images. The *SPP* module can also increase the receptive field so that the most important context features can be extracted without slowing down detection speed. *PANet* module, *Path-Aggregation Network* has feature images of different scales. Through the fusion of different scales, the detailed feature images from the bottom can be directly transmitted to the top for fusion during the process from the bottom to the top to improve the performance of feature extraction and then conduct the operation of concatenation.

(3) Loss Function

The loss function of *YOLOv5* model includes classification loss, location loss and confidence loss. When calculating the classification loss, the model will use the binary cross entropy loss function to calculate the loss of category probability and target confidence score, avoiding the use of *Softmax* function to reduce the computational complexity. See the formula for the specific definition:

$$L_{logistic}(\hat{y}, y) = -ylog\hat{y} - (1-y)\log(1-\hat{y}) \qquad (1)$$

Where, specifically, $y \in \{0,1\}$, $\hat{y} = \sigma(\bar{y}), \sigma(\bar{y}) = \frac{1}{1+e^{-\bar{y}}}, \frac{\partial \sigma}{\partial \bar{y}} = \sigma(1-\sigma)$, $\bar{y} = w \cdot x + b = \sum_j w_j x_{ij} + b, \frac{\partial \bar{y}}{\partial w_j} = x_j$, $x = (x_1, \cdots, x_j, \cdots, x_n)$, $w = (w_1, \cdots, w_j, \cdots, w_n)$, there are *n* values in parameter *x* and *w*, and these values and features in samples are also correspondent.

The model of Yolov5 used *CIoU Loss* to replace the regression Loss, which was used to calculate the Loss of bounding box regression. The penalty term of *CIoU* is based on the *DIoU* penalty term, adding an impact factor αv, which considers to use the length-width ratio of the prediction box to match the length-width ratio of the target box. The specific function formula is as follows:

$$L_{CIoU} = 1 - IoU + \frac{\beta^2(b, b^{gt})}{c^2} + \alpha v \qquad (2)$$

Where, the formula $\alpha = \frac{v}{(1-IoU)+v}$ was used to calculate the parameter of *trade-off*; and the formula $v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2$ was used to measure the consistency of length-width ratio.

(4) Target box regression and cross-grid prediction strategy

Target box regression is mainly used to determine the location of the target box in the image. By regression between the target box and the prior box, the model will obtain the offset between width and height of the true target and the initial target, that is, the relative offset of the center point of the predicted box relative to the top left corner of the corresponding grid. By using the Sigmoid function to process the offset, the model will make the predicted offset within the range of (0,1), and restricts the center point of the bounding box to the current grid, avoiding the runaway gradient, instability, *NAN* loss and eventual complete loss of training caused by the unrestricted width and height.

The formula of target box calculation in *YOLOv5* is:

$$\begin{aligned} b_x &= 2\sigma(t_x) - 0.5 + c_x; \\ b_y &= 2\sigma(t_y) - 0.5 + c_y \\ b_w &= p_w(2\sigma(t_w))^2; \\ b_h &= p_h(2\sigma(t_h))^2 \end{aligned} \qquad (3)$$

Where, t_x, t_y, t_w, t_h are the coordinate transformation coefficient, c_x, c_y are center coordinate of Anchor, p_w, p_h represents the width and height of the target box, $\sigma(\ )$ means the function of sigmoid.

The model of *YOLOv5* adopts a cross-neighborhood grid matching strategy, that is, from the top, bottom, left and right of the current grid to find the two grids closest to the target center point, and then add the current grid for the three grids to match. Through the cross-grid prediction strategy, more positive sample Anchors can be obtained, which can achieve accelerated convergence.

2. Model training

In this paper, the *YOLOv5* model is built under the deep learning framework *PyTorch*, and the target detection model is developed by using the training datasets of Siberian Crane, so as to achieve target classification and location. The training datasets adopted in this experiment mainly includes Train_1, Train_2, Train_3, Train_4, Train_5 and Train_6. The default parameters are used for the hyper-parameters of the model, that is, the weight decay rate is 0.005, the initial learning rate is 0.001, the momentum is 0.9, and the total iteration is 40,000 times. The experiment is conducted based on the Linux system issued by Ubuntu 18.04 LTS. The configuration processor is Intel (R) Core (TM) I7-8700K CPU, the computer memory is 16GB, and the graphics card model is NVIDIA GeForce GTX 2080 TI.

3. Assessment criteria

AP (denoted by A_p), is a general performance evaluation index for predicting the position and category of a single target object. A_p is calculated by precision rate (*P*) and recall rate (*R*), that is, the area enclosed by *P-R* curve is A_p value. *P-R* curve is a two-dimensional curve with precision and recall as vertical and horizontal coordinates. The overall trend is drawn by selecting the accuracy and recall rate corresponding to different thresholds. The higher the accuracy, the lower the recall rate. When the recall reaches 1, it corresponds to the positive sample with the lowest probability score. In this case, the number of positive samples divided by the number of all samples greater than or equal to the threshold value is the lowest accuracy value.

$$P = \frac{T_p}{T_p + F_p} \qquad (4)$$

$$R = \frac{T_p}{T_p + F_N} \qquad (5)$$

Where, T_p is the number of Siberian crane in positive samples of results of detection, which is correctly classified and located; F_p is the number of negative samples labeled as positive samples of Siberian crane, and F_N is the number of positive samples of Siberian crane labeled as negative samples. In this paper, a positive sample is defined as the value of *IOU* greater than 0.8. *mAP*, as an index to measure detection accuracy in target detection, refers to the average *AP* value of multiple verification sets of individuals. Its calculation formula is as follows:

$$mAP = \frac{1}{N}\sum_{n=1}^{N} A_P(n) \qquad (6)$$

III. RESULT AND ANALYSIS

A. Analysis of Siberian crane detection based on YOLOv5

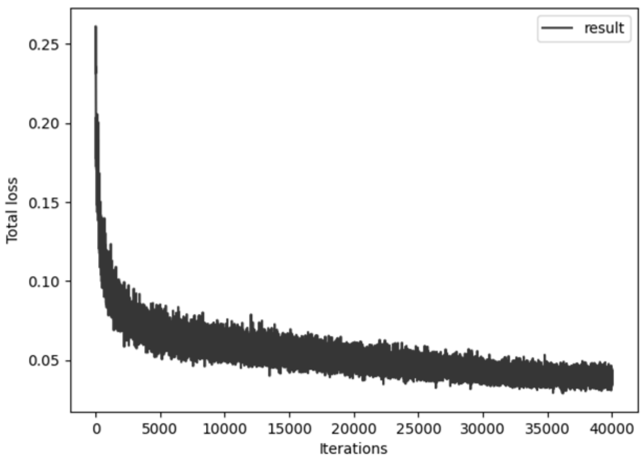

Fig. 4 Change trend of total loss value of the model of *YOLOv5*

As shown in Figure 4, this paper takes the total loss value of 5000 iterations as the interval point to draw the trend line. At the beginning of the iteration, the total loss value of the model is 0.2354. With the increase of the number of iterations, the total loss value shows a downward trend and the *mAP* keeps increasing. When the number of iterations is greater than 5000, the decline rate of the total loss value gradually decreases, and the model gradually converges until the number of iterations is greater than 35,000, which tends to be flat. When the number of iterations is equal to 40,000 times, the total loss value is 0.0357, and the *mAP* can reach 81.3%. In order to select a better number of iterations, models with iterations of 30000, 35000 and 40000 times were tested in this paper. The test results are shown in Figure 5.Through comparison, it is found that the model with 40000 iterations detects the highest detection rate of Siberian crane in the same image. In terms of running time, the time needed for 30000 iterations is 87 h, the time needed for 35,000 iterations is 102 h, and the time needed for 40000 iterations is 116 h. The time required for 40000 iterations is relatively long, but a relatively good detection effect can be obtained. Therefore, 40000 times is selected as the optimal number of iterations in this paper.

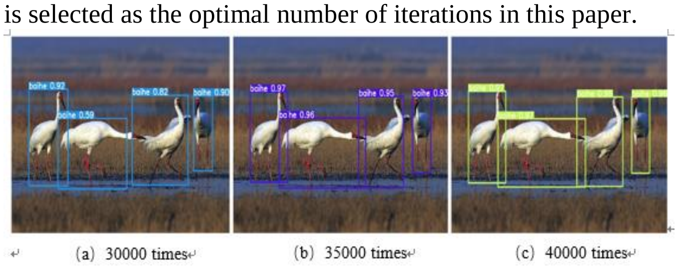

Fig.5 Detection of *YOLOv5* at different times of iterations

B. Comparison of detection accuracy of Siberian crane under different training datasets

It can be seen from Table 2 that the *mAP* values of the six training datasets are Train_4 > Train_2 > Train_1 > Train_3 in turn. Among them, Train_5 and Train_6 are not tested because of the low accuracy of the training dataset panoramic images. he *mAP* value of Train_4 was 90.9%, which was much higher than other training dataset, indicating that the model detection accuracy of the training set mixed with ordinary field photos and network photos was much higher than that of the training set from a single source image, and it was suitable for subsequent scientific research. The low detection accuracy of Train_3 is mainly due to the fact that the Siberian crane is far away from the camera due to its own alertness in the field when we are taking the photos, so the target is too small in the picture and is not easy to be detected. At the same time, Mosaic data enhancement method was adopted in *YOLOv5*, which further reduced the originally small target body, resulting in an accuracy of almost zero and being recognized as the background.

Table 2 Influence of different training datasets on detection effect of Siberian crane

Index	Training dataset	*Precision* (%)	*Recall* (%)	*mAP* (%)
1	Train_1	91.5	41.2	45.4
2	Train_2	89.4	80.3	81.3
3	Train_3	4.6	5.2	0.7
4	Train_4	98.8	88.6	90.9
5	Train_5	----	----	----
6	Train_6	----	----	----

C. Analysis of detection effect of YOLOv5 in complex environments

Fig6. Detection results of partial Siberian crane images

As can be seen from Figure 6, in a complex environmental background, the detection results of Siberian crane based on *YOLOv5* model performed well. There are small multi-target objects in the image and some of them are occluding each other, but the detection performance of the *YOLOv5* model is more than 90%. The *YOLOv5* model has excellent detection performance when the image is covered by wire and the color of the target body is similar to that of the background environment. When the Siberian crane is flying, falling, foraging, playing and other different activities, the model can still carry out effective detection. Therefore, this model can effectively detect individual cranes or clusters under complex environmental background.

IV. Conclusion

Based on the *YOLOv5* model, this paper uses images from three different sources (Internet, ordinary, panoramic field images) to build six different datasets. By comparing the detection metrics (*Precision*, *recall* and *mAP*) of the *YOLOv5* model with 6 different datasets and iteration times , the results show that the *mAP* values of the six training datasets are Train_4 > Train_2 > Train_1 > Train_3 in turn. The *mAP* value of Train_4 was biggest, reaching to 90.9%, which was much higher than other training dataset, and indicate that the model detection accuracy of the training dataset mixed with ordinary field photos and network photos was much higher than that of the training dataset from a single source image; when the iteration times reach to 40000times, *YOLOv5* model can completely converge, and *mAP* value reached to 81.3%, total loss value 0.0357； According to the result of the existing model test, we found that the model can also have an effectively performance of detection in the complex environments, such as pictures exist the problems of multi-objective small objects and occlusions, similarity in color between the object and background, and Siberian cranes are in different activities such as flying, falling, foraging, playing etc.

Acknowledgment

The paper was supported by the Fundamental Research Funds of IFRIT(CAFYBB2019SZ004), the National Natural Science Foundation of China (32071681) and the National Key Research and Development Program of China (2017YFD0600905). The paper's corresponding author is Huaiqing Zhang.

References

[1] Zou L, Hu B, Qi S, et al. Spatiotemporal Variation of Siberian Crane Habitats and the Response to Water Level in Poyang Lake Wetland, China[J]. Remote Sensing, 2021, 13(1): 140.

[2] Wen D, Hu Y, Xiong Z, et al. Potential Suitable Habitat Distribution and Conservation Strategy for the Siberian Crane (Grus leucogeranus) at Spring Stopover Sites in Northeastern China[J]. Polish Journal of Environmental Studies, 2020, 29(5).

[3] Jiang H, He C, Luo W, et al. Hydrological restoration and water resource management of Siberian Crane (Grus leucogeranus) stopover wetlands[J]. Water, 2018, 10(12): 1714.

[4] Burnham K P, Anderson D R, Laake J L. Estimation of density from line transect sampling of biological populations[J]. Wildlife monographs, 1980 (72): 3-202.

[5] Merz G. Counting elephants (Loxodonta africana cyclotis) in tropical rain forests with particular reference to the Tai National Park, Ivory Coast[J]. African Journal of Ecology, 1986, 24(2): 61-68.

[6] Groom G, Stjernholm M, Nielsen R D, et al. Remote sensing image data and automated analysis to describe marine bird distributions and abundances[J]. Ecological Informatics, 2013, 14: 2-8.

[7] Groom G, Krag Petersen I, Anderson M D, et al. Using object-based analysis of image data to count birds: mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa[J]. International Journal of Remote Sensing, 2011, 32(16): 4611-4639.

[8] Linchant J, Lisein J, Semeki J, et al. Are unmanned aircraft systems (UAS s) the future of wildlife monitoring? A review of accomplishments and challenges[J]. Mammal Review, 2015, 45(4): 239-252.

[9] Dunham K M. Trends in populations of elephant and other large herbivores in Gonarezhou National Park, Zimbabwe, as revealed by sample aerial surveys[J]. African Journal of Ecology, 2012, 50(4): 476-488.

[10] Liu C C, Chen Y H, Wen H L. Supporting the annual international black-faced spoonbill census with a low-cost unmanned aerial vehicle[J]. Ecological informatics, 2015, 30: 170-178.

[11] Chabot D, Bird D M. Evaluation of an off-the-shelf unmanned aircraft system for surveying flocks of geese[J]. Waterbirds, 2012, 35(1): 170-174.

[12] Grenzdörffer G J. UAS-based automatic bird count of a common gull colony[J]. International archives of the photogrammetry, Remote sensing and spatial information sciences, 2013, 1: W2.

[13] Chabot D, Bird D M. Evaluation of an off-the-shelf unmanned aircraft system for surveying flocks of geese[J]. Waterbirds, 2012, 35(1): 170-174.

[14] Hong S J, Han Y, Kim S Y, et al. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery[J]. Sensors, 2019, 19(7): 1651.

[15] Ammour N, Alhichri H, Bazi Y, et al. Deep learning approach for car detection in UAV imagery[J]. Remote Sensing, 2017, 9(4): 312.

[16] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[17] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[18] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[19] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. arXiv preprint arXiv:1506.01497, 2015.

[20] 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 and machine intelligence, 2016, 39(6): 1137-1149.

[21] 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: 779-788.

[22] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[23] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[24] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[25] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[26] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[28] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[29] Li S, Deng M, Lee J, et al. Imaging through glass diffusers using densely connected convolutional networks[J]. Optica, 2018, 5(7): 803-813.

[30] Hui F, Xu B, Huang H, et al. Modelling spatial - temporal change of Poyang Lake using multitemporal Landsat imagery[J]. International Journal of Remote Sensing, 2008, 29(20): 5767-5784.

[31] Li Y, Qian F, Silbernagel J, et al. Community structure, abundance variation and population trends of waterbirds in relation to water level fluctuation in Poyang Lake[J]. Journal of Great Lakes Research, 2019, 45(5): 976-985.