International Journal of Science and Engineering Applications
Volume 13-Issue 03, 26 - 29, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1303.1006

Ostrich Detection based on Deep Learning

Tingting Wang
School of Electric Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: With the modernization and intelligence of agricultural production, the demand for animal management and monitoring is also increasing. Ostriches are an important livestock resource that requires effective monitoring and identification in breeding and management. This article proposes an improved YOLOv5s based ostrich detection model by establishing a self-made ostrich detection dataset and manually annotating it. The model introduces a lightweight network GhostNetV2, integrates position attention mechanism, uses collaborative coordinate convolution module, and replaces the loss function with MPDIoU to achieve feature extraction, reduce the number of parameters, and computational complexity Comprehensive optimization in enhancing feature fusion capability and improving the quality of target detection results. At the same time, the self-made dataset also ensures the effectiveness of the model in practical application scenarios, providing a solid theoretical foundation for subsequent ostrich breeding management, early disease warning, and agricultural intelligent development.

Keywords: Image recognition; Deep learning; Ostrich identification; object detection; image processing

1. INTRODUCTION

Nowadays, intelligent breeding of livestock and poultry has gradually become a research hotspot [1]. This is not only because intelligent breeding can improve the growth efficiency of livestock and poultry, but also because it can help us better understand the growth process and disease status of animals. Ostriches, as a special type of poultry, have very high economic value and special use. At present, in the process of ostrich breeding, breeders usually monitor the health status and behavioral changes of ostriches through visual observation. They need to keep long-term records of ostrich activity and diet in order to detect abnormal situations in a timely manner. However, this approach requires a significant amount of manpower, time, and effort, and is easily influenced by human factors, resulting in poor detection results and untimely decision-making. To solve this problem, we can combine computer vision technology with artificial intelligence technology and apply it to the ostrich breeding industry. By using these technologies, we can achieve automated monitoring and intelligent management, improve breeding efficiency and the health status of ostriches. Specifically, we can use computer vision technology to detect ostriches, and then use artificial intelligence technology to track and analyze the trajectory and behavior of the detected targets.

In terms of object detection, we can use deep learning algorithms to train models that can automatically recognize and position ostriches. By training a large amount of annotated data, the model can gradually improve accuracy and ultimately achieve high-precision object detection. This step is the foundation for subsequent individual trajectory tracking and pathological detection, so it is very important. Through the application of such technology, we can more timely detect the abnormal behavior and disease symptoms of ostriches, and take corresponding measures for intervention and treatment. This can not only improve the growth efficiency and economic value of ostriches, but also provide technical support and guarantee for the sustainable development of ostrich breeding industry.

2. MATERIAL AND METHODS

2.1 Data Acquisition

Due to limited research on ostrich target detection, it is difficult to obtain publicly available ostrich target detection datasets. In this study, a self-made ostrich target detection dataset was used, with the video sourced from an iPhone camera and a resolution of 1920x1080. The video image was saved in. mp4 format. Python scripts were used to extract ostrich activity videos from different time periods, convert them into images, and extract frames from each video at certain intervals, After filtering, a total of 685 images were retained, with a unified numbering format of "00X. jpg", where X represents the image number.

2.2 Data Preprocessing

Use the LabelImg tool to label the target with the minimum bounding box for the collected 685 images. Save the annotated results in the txt format used by the YOLO model, and annotate all ostrich targets in the image as much as possible while being distinguishable by the human eye. For ambiguous ostrich targets, it is possible to choose not to label them to avoid treating unlabeled targets as negative samples, thereby reducing the algorithm's ability to distinguish between positive and negative samples. In real ostrich breeding centers, there are situations such as railing obstruction, incomplete exposure of ostrich targets, and adhesion. For these different situations, use appropriately sized rectangular boxes for annotation to ensure the accuracy and consistency of the annotation results.

In order to improve the richness and robustness of the data, data augmentation techniques were adopted to process the collected dataset. Specifically, operations such as rotation, horizontal mirror flipping, and Gaussian blur transformation were performed on each image and its annotation boxes to generate more training samples. In addition, operations such as random brightness adjustment and pixel value multiplier change were also used to increase the diversity and randomness of the data. The images were randomly flipped between -16 ° and 16 °, and the brightness of the images and annotation boxes was randomly adjusted. Each pixel was randomly multiplied by a value between 1.2 and 1.5. Through data augmentation techniques, 3 new images were generated

International Journal of Science and Engineering Applications
Volume 13-Issue 03, 26 - 29, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1303.1006

for each original image, resulting in a total of 2740 enhanced image datasets. The enhanced images not only have the diversity of the original images, but also introduce new randomness, Figure 1 shows an example of the enhanced image:

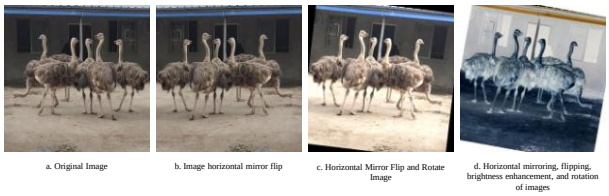

Figure. 1 Images after dataset enhancement

For the expanded ostrich detection dataset, it was randomly divided into training, validation, and testing sets in a 6:2:2 ratio. The specific dataset partitioning is shown in Table 1.

Table 1. Dataset partitioning situation

Dataset	Number of images
Training set	1644
Validation set	548
Test set	548

2.3 YOLOv5s Algorithm

YOLOv5 is a fast, flexible, and end-to-end computer vision algorithm. Compared to other versions of the YOLO [2] series, YOLOv5s is the smallest version of YOLOv5, with the least layers and computational complexity, resulting in the smallest model size. It consists of four parts: input layer, backbone network layer, feature fusion layer, and output layer [3]. The specific structure diagram is shown in Figure 2.

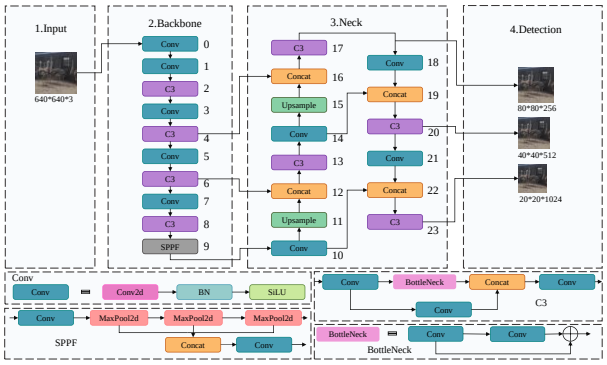

Figure. 2 YOLOv5s

(1) Input: Utilizing Mosaic data augmentation to increase the diversity and richness of the dataset, this method randomly combines four images together to form a large image, and then uses the large image containing four different images as the training set image for data augmentation operations such as scaling, cropping, flipping, and color gamut changes to enrich the dataset image. Then, the anchor box parameters of the most suitable image are automatically calculated, Without the need for manual settings, it reduces the impact of samples and improves detection accuracy.

(2) Backbone: The CSPDarknet53 network is used to extract the feature information of the target to be detected through a deep convolutional neural network, and the results are input into the neck network. Focus is the first convolutional layer in the network, which down-samples the feature map obtained from the input layer, extracts important information, and achieves down-sampling and compression of the feature layer. The CSP structure processes a portion of the input feature map through a sub network, while the other portion enters the convolution operation of the next layer, connecting these two parts of the feature map as input to the next layer.

(3) Neck: Adopting an FPN+PAN structure [4], it achieves feature fusion and extraction reinforcement of the network model. FPN integrates features in a bottom-up manner, combining feature information from different scales to enable the model to simultaneously obtain information from different scales; PAN aggregates feature maps from different paths to obtain richer contextual information, enhancing the model's fusion ability for detecting targets.

(4) Detection: This layer consists of three different scale prediction layers, aimed at detecting large-scale, mesoscale, and small-scale targets. During this process, anchor boxes are used to predict the position information and size of the target bounding box, output and display the detected target category and its probability.

2.4 Improved YOLOv5s Algorithm

Figure 3 shows the structure diagram of an improved ostrich target detection algorithm based on the original YOLOv5s.

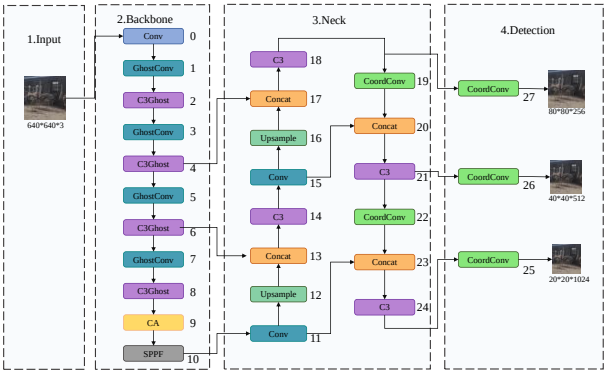

Figure. 3 Improved YOLOv5s

The improvements are as follows:

(1) Introducing a lightweight GhostNetV2 network into the backbone network and integrating positional attention mechanism to replace the original backbone network of the network model, reducing the number of model parameters while enhancing feature extraction ability.

(2) Introducing Coordinated Convolutional in the neck and detection head enables the network to have spatial awareness and better sense the positional information in the feature map.

(3) In the algorithm model, the original CIoU loss function is replaced with MPDIoU, which can directly optimize the IoU in the object detection task, making the model more accurate in locating and capturing targets, and improving the quality of the model's object detection results.

The improved ostrich target detection model not only reduces the number of network parameters and computational complexity, but also ensures the detection accuracy of the model. This enables the model to meet the requirements of ostrich target detection in complex outdoor backgrounds while ensuring performance, and accurately complete the detection task.

2.4.1 GhostNetV2

At present, ostrich detection has the problem of a large number of model parameters, which leads to a large

International Journal of Science and Engineering Applications
Volume 13-Issue 03, 26 - 29, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1303.1006

consumption of memory and affects the detection speed of the model. A common method is to use lightweight network structures to reduce the number of model parameters and computational complexity, which can improve model detection speed. By replacing the backbone network with the lightweight network model GhostNetV2 [5], the parameter and computational complexity of the model can be further reduced. Replacing the C3 structure with multiple GhostNetV2 structures can reduce the parameter size of the entire model. There is a large amount of redundant information in deep learning networks, which increases the complexity and computational complexity of the model, although it helps to maintain high accuracy. In order to reduce the number of parameters while avoiding negative impacts on model accuracy, the GhostNetV2 network is adopted. GhostNetV2 is an improved version of the GhostNet network, which reduces the number of parameters by calculating the map of redundant information without affecting the accuracy of the model.

The overall structure of GhostNetV2 is composed of Ghost Bottlenecks and some structures. It adopts a reverse residual bottleneck that includes two Ghost modules. The first module generates extended features with more channels, while the second module obtains output features by reducing the number of channels. The DFC attention module runs in parallel with the first Ghost module to enhance the extension function, and the second Ghost module receives enhanced features, And generate output features. GhostNetV2 bottleneck is shown in Figure 4:

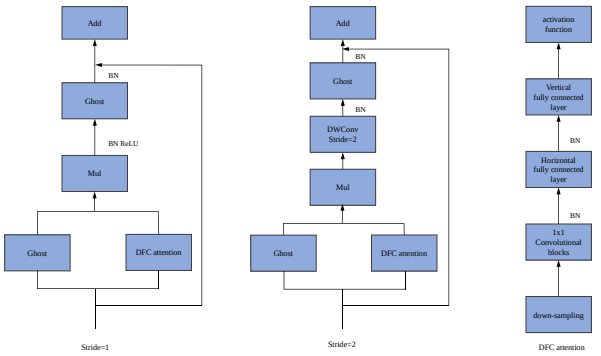

Figure. 4 Ghostnetv2 bottleneck module

2.4.2 CoordAttention

The attention mechanism enables neural networks to selectively focus their attention on key information, avoiding information redundancy and effectively improving the efficiency and accuracy of object detection. The application of attention modules has been widely recognized, including spatial attention modules, channel attention modules [6], fused channel spatial attention modules [7], and positional attention mechanisms [8]. The channel attention module can adjust the weight of each channel to enhance the attention to important features and improve the performance of the model [9]. On the basis of SE, CBAM further reduces the dimensionality of feature maps and uses large-scale convolution to encode spatial information, achieving multi-dimensional feature optimization. However, although CBAM considers channel information and position information of features, it separates them. The position attention module can effectively capture position information and channel relationships, embed position information into feature maps, and generate coordinate attention to capture the characteristics of different positions in the image, making the network more accurate in locating objects of interest. Figure 5 shows the structure of CA:

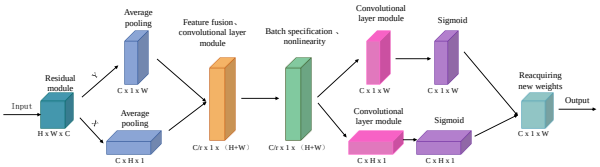

Figure. 5 CA attention module

2.4.3 Coordinated Convolution

Introducing coordinated coordinate convolution, the input coordinate information is integrated into the feature map of the neural network layer. By perceiving spatial information, the model can better perceive the position information in the feature map. The CoordConv module [10], as shown in Figure 6, adds two coordinate channels after the extracted feature map, representing the i and j coordinates of the input feature map, so that each pixel not only has a color channel (RGB), but also adds two coordinate channels to represent the position of the pixel in the feature map. If the coordinate channel learns a certain amount of information, then CoordConv has translation dependency. When there is no learning information, CoordConv is equivalent to traditional convolution having translation invariance.

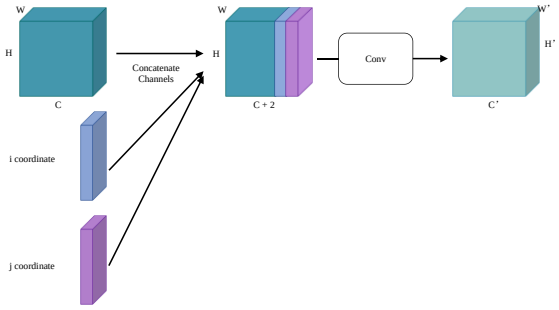

Figure. 6 Coordinate convolution module

3. EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental Environment and Parameter Settings

The operating environment for this experiment is Windows 10 Professional Edition, with a hardware environment of 12th Gen Intel (R) Core (TM) i5-12400F CPU processor, NVIDIA GeForce RTX 3060 graphics card, Python version 2021.1.3 x64 software programming language, PyTorch deep learning framework version 1.13.1, CUDA version 11.7, and CUDNN version 8.4.0.

During the training process, the input of the network is 640x640 pixels, the initial learning rate is adjusted to 0.01, the momentum factor is set to 0.937, and the optimizer uses a stochastic gradient descent (SGD) optimizer with a total of 300 training iterations. The batch size is set to 8.

3.2 Experimental Results

Figure 7 shows the comparison of ostrich detection results between the original YOLOv5s and the improved YOLOv5s. Among them, Figure a is the manually annotated image, Figure b is the result image of YOLOv5s detection, and Figure c is the improved detection result image of YOLOv5s. From the comparison effect image, it can be seen that the improved detection algorithm has improved confidence in

outdoor environments such as incomplete ostrich exposure and adhesion, proving the effectiveness of the algorithm.

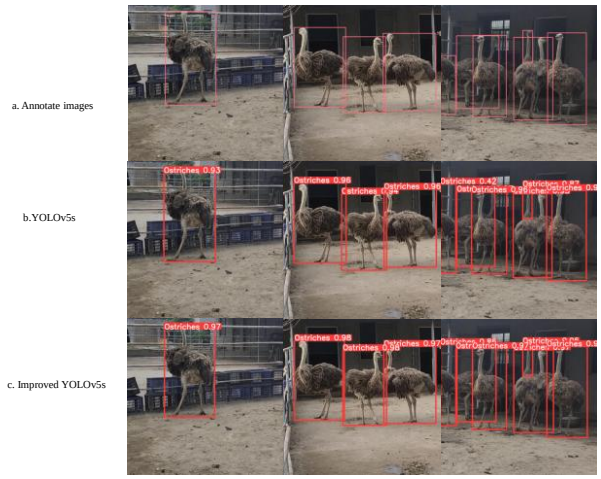

Figure. 7 Detection comparison chart

4. CONCLUSION

This article proposes an ostrich detection model based on improved YOLOv5s, which uses the lightweight network GhostNetV2 as the backbone feature extraction network, and adds a CA position attention mechanism to further enhance feature representation and model performance. Meanwhile, Path Aggregation Network (PANet) and CoordConv were used in the neck network and detection head to better utilize coordinate information to enhance feature representation and model performance. Finally, the original CIoU loss function was replaced with an MPDIoU loss function to better measure the gap between the predicted box and the true box, optimizing the performance of object detection. The improved YOLOv5s model can meet the detection needs of high accuracy and fast detection speed, laying the foundation for ostrich behavior monitoring and early disease warning. This method can also be extended to the detection and recognition of other poultry, providing a theoretical basis for individual tracking and disease monitoring in actual breeding farms.

5. REFERENCES

[1] Zhu Jing, Li Tingting, Shi Shourong, et al. Development status, existing problems, and countermeasures of pigeon, quail, and special poultry industries in China [J]. Chinese Journal of Animal Husbandry, 2024 (001): 060

[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition.IEEE, 2016.DOI:10.1109/CVPR.2016.91.

[3] Liu Guang, Hu Guoyu, Gulibahar Tuohuti et al. Detection of grape leaf diseases and pests based on improved YOLOv3 [J/OL]. Microelectronics and Computer Science, 2023, (02):110-119

[4] Xu Xiang, Cai Maoguo, Tang Jianlan. Research on Target Recognition Based on Improved YOLOv4 [J]. Information Technology, 2022, 46 (12): 107-111

[5] Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., & Wang, Y. (2022). GhostNetV2: Enhance Cheap Operation with Long-Range Attention. ArXiv, abs/2211.12905.

[6] Hu J, Shen L, Sun G,et al.Squeeze-and-Excitation Networks[C]//IEEE.IEEE, 2017.DOI:10.1109/TPAMI.2019.2913372.

[7] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module[J]. Springer, Cham, 2018.DOI:10.1007/978-3-030-01234-2_1.

[8] Q. Hou, D. Zhou and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 13708-13717, DOI: 10.1109/CVPR46437.2021.01350.

[9] Han Gujing; He Min; Lei Yuhang; Zhang Min; Zhao Liu; Qin Liang. Research on Image Segmentation of Transmission Line Insulators Based on Improved U-Net [J]. Smart Power, 2022, 50 (03): 93-99

[10] Liu R, Lehman J, Molino P, et al. An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution[J]. 2018.DOI:10.48550/arXiv.1807.03247.