International Journal of Science and Engineering Applications
Volume 13-Issue 10, 132 – 137, 2024, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1310.1027

Lightweight Horse Detection Model based on Improved YOLOv8

Haijun Wu
School of Electric Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: In the field of agriculture, the accurate detection of horses can not only improve the efficiency of ranch management, but also provide important data support for horse health monitoring and behavior analysis, and help promote the intelligent and sustainable development of agriculture. Aiming at the intelligent needs of modern agriculture and animal husbandry, this study proposes an improved horse detection model for YOLOv8. By introducing the GSConv module in the backbone network to strengthen the feature extraction capability, and using the VoV-GSCSP module to replace the C2f module in the neck network, the detection accuracy and computational efficiency are significantly improved. The experimental results show that combining the GSConv and VoV-GSCSP modules increases the precision of the model by 0.9%, the recall by 1.9%, and the mean average precision (mAP) by 1.9%, while the number of parameters and the computational effort are reduced by 5% and 11%, respectively. The improved model demonstrated higher accuracy and efficiency in the horse detection task and reduced false and missed detections, proving its potential application value in horse detection, which can provide a reference for intelligent horse management in the future.

Keywords:YOLOv8; Horse detection; lightweighting; GSConv; VoV-GSCSP

1. INTRODUCTION

In modern agriculture and animal husbandry, precise monitoring and management of animals is crucial for production efficiency and economic benefits. With the increasing demand for precision agriculture, the use of advanced technology to achieve efficient detection of specific animals has become one of the core issues to promote the intelligent development of agriculture. Target detection, as a key technology of computer vision, has demonstrated extensive value in practical applications. Through this technology, the agricultural field is able to realize the automatic identification of animals, which in turn improves the level of production intelligence, reduces human input, and effectively reduces the waste of resources [1].

Among many animals, horse detection has important research significance due to its unique role in agricultural production and ecosystem. Horses occupy a key position in the fields of rangeland management, agricultural economics, sports science, and ecological conservation. High-precision target detection technology can realize real-time identification of horses and accurately obtain key information such as their health status and behavioral characteristics. This type of automated detection technology can not only significantly improve the efficiency of pasture management, but also provide a reliable scientific basis for horse health monitoring, thus promoting the sustainable development of agriculture. In addition, the automated detection technology of horses provides valuable data support for agricultural research and promotes technological innovation and progress in related fields.

In this study, a horse detection model based on the improved YOLOv8 is proposed in this context. By introducing the GSConv module in the backbone network of the model, replacing part of the traditional convolutional layer (Conv) to enhance the feature extraction capability, and replacing all the C2f modules in the neck network of the model with the novel computational block VoV-GSCSP, we ensure that the network improves the detection accuracy while maintaining the light weight [2]. This optimization scheme aims to provide a reference for the development of smart agriculture.

2. YOLOv8 Algorithm

YOLOv8 [3] is a target detection algorithm released by Ultralytics in 2023 that further pushes the technological boundaries of speed, accuracy, and user-friendliness in the field of deep learning.YOLO stands for You Only Look Once, and its innovation is to predict all bounding boxes simultaneously through a single forward propagation, thus significantly improving the efficiency and real-time processing capability of the algorithm. YOLO stands for "You Only Look Once," and its innovation is to predict all bounding boxes simultaneously through a single forward propagation, thus significantly improving the efficiency and real-time processing capability of the algorithm. In contrast to many other target detection methods that require multiple stages to complete target identification and localization, YOLOv8 extends the popular YOLOv5 architecture and provides performance improvements in several key areas. One of its most significant improvements is the introduction of anchorless detection, which simplifies Non-Maximum Suppression (NMS) during post-processing and further increases detection speed.YOLOv8 is able to perform object recognition and localization in images and videos with superior speed and accuracy, while simultaneously handling multi-tasks such as image classification and target detection. The network structure of YOLOv8 is shown in Fig. 1, which consists of four main parts: the input layer (Input), the backbone network (Backbone), the neck network (Neck), and the detection head (Head).

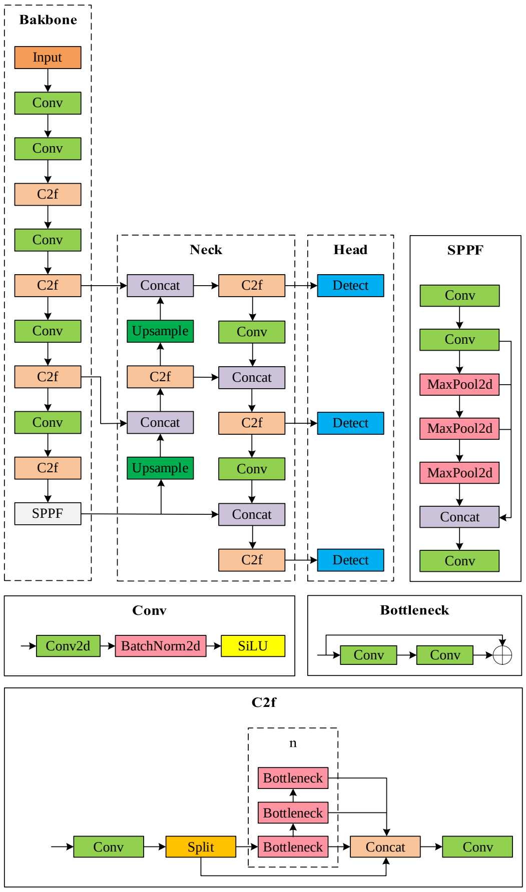

Fig. 1 Structure of YOLOv8 model network

In the Input layer, the model first adjusts the input color image to a size and format suitable for processing. This process includes pre-processing steps such as scaling and normalization of the image to ensure that the image data meets the requirements of model training and inference. These preprocessing operations play an important role in improving the stability and overall performance of the model.

The main task of Backbone is to perform feature extraction, which is operated through a series of convolutional and anti-convolutional layers, and combines residual connectivity with bottleneck structures to reduce network size and enhance performance. The C2f module is used as the base unit for this part. Compared to the C3 module in YOLOv5, the C2f module has a reduced number of parameters, as well as stronger feature extraction capabilities. Specifically, the C2f module reduces redundant parameters through a more optimized structural design, thus improving computational efficiency. In addition, Backbone contains a spatial pyramid pooling layer (SPPF), which is capable of adaptively fusing information from different scales to further enhance the performance of target detection.

Neck is responsible for fusing multi-scale features. It adopts the PAN-FAN structure, which further enhances the feature representation capability of the model by combining feature maps from different layers of Backbone.

Finally, Head is responsible for processing the feature maps from Neck and generating the final detection results. Its main function is to convert the feature maps into specific information required for target detection, such as category, location and confidence.

3. Improved Methods

Deep learning-based detectors have been dominating the field of target detection since the inception of convolutional neural networks (RCNN). However, deep neural networks (DNNs) usually contain a large number of computational parameters, leading to their consumption of large computational resources in practical applications. In order to reduce the computational cost while maintaining the detection accuracy, the model is optimized in this study. In Backbone, the GSConv module is used instead of the convolutional layer (Conv) in the sixth layer of Backbone, which reduces the complexity of the model and maintains the accuracy. In addition, the Neck of the model is designed to be lightweight, which significantly reduces the computational overhead while ensuring the horse detection accuracy. The improved model network structure is shown in Fig. 2.

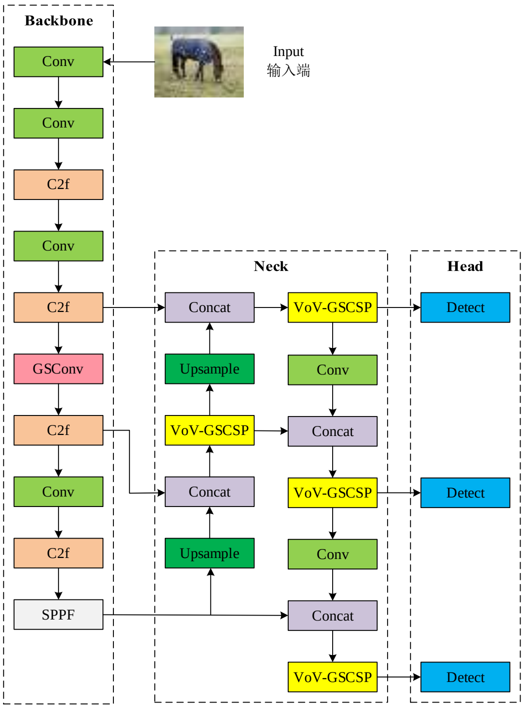

Fig. 2 Improved YOLOv8 model network structure

3.1 GSConv Module

Many lightweight models are constructed from start to finish using depth-separable convolution (DSC) to reduce the number of parameters (params) and the number of floating point operations per second (GFLOPs). Although DSC is effective in lightweighting, its drawback is equally obvious: the channel information of the input image is separated during the computation process, resulting in a significantly lower feature representation capability of DSC than that of standard convolution (SC). This shortcoming is further amplified in the backbone network, where the features generated by simply shuffling the output channels of DSC still exhibit “deep separation” in both classification and detection tasks.

In order to understand the shortcomings of DSC in feature representation, this study replaces GSConv with Conv in the sixth layer of Backbone of the model, and the structure of the GSConv module is shown in Fig. 3. The computational process of GSConv is as follows: first, the input feature map is processed using a standard convolutional kernel to maintain the information exchange between channels. Next, the input feature map is processed using DSC to reduce the computation. Then the feature maps generated by SC and DSC are reorganized to mix the channel information to enhance the feature representation capability. Finally, the reorganized feature map is the final output of GSConv.

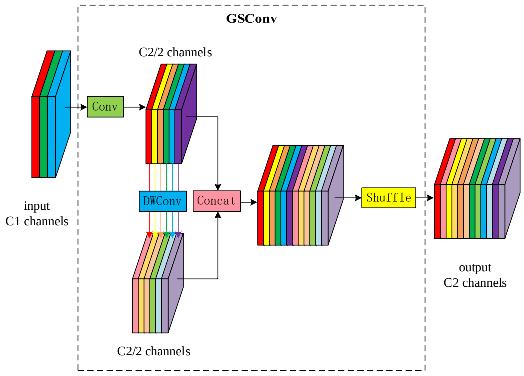

Fig. 3 Structure of GSConv module

3.2 VoV-GSCSP Module

VoV-GSCSP is a novel computational module that enhances the efficiency and accuracy of lightweight convolutional neural networks (CNNs).Built on GSConv, VoV-GSCSP can significantly reduce computational complexity and inference time through efficient feature fusion and utilization strategies. The module uses GSConv as the core and utilizes GSConv for feature reorganization to enhance feature representation.VoV-GSCSP reduces redundant computation and improves feature utilization efficiency through cross-stage feature fusion, which enables the network to share and reuse features across different stages, thus reducing the network computational burden.VoV-GSCSP also employs a partial-network design, which uses only a a portion of the network to process specific inputs rather than the entire network, an approach that further improves the overall efficiency of the network. In this study, the VoV-GSCSP module replaces all the C2f modules in the model Neck to achieve lower computational cost while maintaining accuracy, and the VoV-GSCSP module structure is shown in Fig. 4.

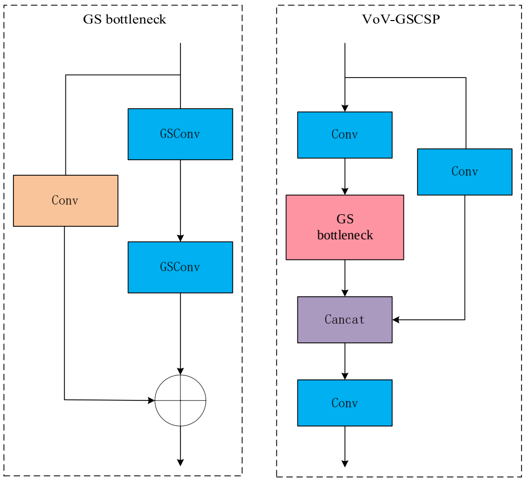

Fig. 4 Structure of VoV-GSCSP module

4. Experiments Results

4.1 Horse Dataset

In this study, a total of 7073 horse images were acquired from Google's Open Images Dataset v7 public dataset via Python scripts. Subsequently, the dataset was labeled and formatted using Roboflow, an open source image annotation platform. The horse dataset was divided according to the ratio of 7:2:1, where 4951 images were used for the training set, 1414 images for the validation set, and 708 images for the test set.

4.2 Experimental environment and Parameter settings

The hardware configuration used in this study includes Intel Core i7-12650H 2.30 GHz CPU, NVIDIA GeForce RTX 4060 8 GB GPU, and 16 GB RAM.For the software environment, Windows 11 operating system is used, Pytorch 2.0.1 deep learning framework, CUDA version 11.8 and Python 3.8 programming language, and the development tool is PyCharm IDE. The training parameters are set as follows: the image size of the input model is 640×640 pixels, the training batch size is 16, the number of training rounds (epochs) is set to 150, the initial learning rate is 0.01, the momentum is 0.937, the SGD optimizer is used, and the rest of the parameters use the default values.

4.3 Evaluation Metrics

In this study, precision (P), recall (R) and mean average precision mean (mAP@0.5) were used to evaluate the accuracy of the model for horse detection, where the threshold of mAP was set to 0.5. The complexity of the model was evaluated by the number of parameters (Params) and the number of floating point operations per second (GFLOPs) [4].

4.4 Ablation Experiments

To assess the impact of introducing the GSConv module and the VoV-GSCSP

module on the performance of the horse detection model, this study conducted ablation experiments to assess the impact of individual network structure branches in the enhanced YOLOv8 algorithm on its overall performance. The performance of the YOLOv8 model and its optimized variants using different enhancement combinations were compared and analyzed using the test data set. The experimental results for the test set are shown in Table 1.

Table 1 Results of ablation experiments

Note: (1) represents YOLOv8; (2) represents YOLOv8+GSConv; (3) represents YOLOv8+GSConv+VoV-GSCSP

Models	P	R	mAP	Params	GFLOPs
(1)	85.3	72	80.9	3M	8.2G
(2)	87.3	71.4	81.5	2.97M	8G
(3)	86.2	73.9	82.8	2.85M	7.3G

Table 1 shows that the introduction of the GSConv module improves the precision and mAP of the model by 2% and 0.6%, respectively, while the number of parameters and computation are reduced by 1% and 2.5%, respectively. After combining both GSConv and VoV-GSCSP modules, the model still improves the precision rate by 0.9%, recall and mAP by 1.9%, while the number of parameters and computation are reduced by 5% and 11%, respectively, compared with the original model. The improved model outperforms the original model in terms of detection accuracy and effectively reduces the phenomena of misdetection and omission, while there is a significant decrease in the number of parameters and computational cost.

4.5 Comparative Experiments

In order to verify the performance advantages of the improved YOLOv8 model in the horse detection task, we selected other models of the YOLO series for comparison experiments, and the results are summarized in Table 2. As shown in the data in the table, the improved YOLOv8 model shows better performance in key metrics, such as the recall rate and the mAP, and the amount of computation is significantly reduced. Compared with the YOLOv5 model, the precision, recall and mAP are improved by 1.2%, 0.6% and 2.1%, respectively, while the number of parameters and computation are reduced by 4.16M and 8.5G, respectively.Compared with the YOLOv10 [5] model, the recall and mAP are improved by 5.8% and 3.6%, respectively, and the computation is reduced by 0.9G.Therefore, the improved YOLOv8 model improves the precision, recall and mAP of horse detection while effectively reducing the consumption of computational resources, and exhibits superior performance over other YOLO series models.

Table 2 Comparative experimental results

Models	P	R	mAP	Params	GFOPs
YOLOv5	85	73.3	80.7	7.01M	15.8G
YOLOv8	85.3	72	80.9	3M	8.2G
YOLOv10	87.4	68.1	79.2	2.69M	8.2G
Ours	86.2	73.9	82.8	2.85M	7.3G

5. CONCLUSION

In this study, a lightweight horse detection model based on improved YOLOv8 is proposed, and its main conclusions are as follows:

(1) The introduction of GSConv and VoV-GSCSP modules in the YOLOv8 model significantly improves the performance of horse detection. The improved model shows excellent improvement in key metrics such as precision, recall and mAP, and significantly reduces the number of parameters and computation.

(2) Compared with other models in the YOLO series, the improved YOLOv8 model not only has advantages in detection accuracy, but also significantly reduces the consumption of computational resources. Experiments show that the improved YOLOv8 model possesses higher accuracy and efficiency in the horse detection task, effectively reduces the phenomenon of misdetection and omission, and verifies its potential application in horse detection, which can provide a reference for the development of intelligent agriculture.

6. REFERENCES

[1] Alif M A R, Hussain M. YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain[J]. arXiv preprint arXiv:2406.10139, 2024.

[2] Li H, Li J, Wei H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022.

[3] Terven J, Córdova-Esparza D M, Romero-González J A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

[4] M.F. Deng, J.J. Gong, P.Y. Zheng, et al. A lightweight target detection method based on improved YOLOX for group-reared hogs[J]. Journal of Agricultural Machinery, 2023, 54(11): 277-285.

[5] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.

[6] DING Qi-An, LIU Long-Shen, CHEN Jia, et al. Target detection of lactating piglets based on Jetson Nano+ YOLO v5[J]. Journal of Agricultural Machinery, 2022, 53(3): 277-284.

[7] Wang X, Liu J. Vegetable disease detection using an improved YOLOv8 algorithm in the greenhouse plant environment[J]. Scientific Reports, 2024, 14(1): 4261.

[8] TAN Hou-Sen, MA Wen-Hong, TIAN Yuan, et al. An improved YOLOv8n-based target detection method for balsam pear[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(11).