International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

Ostrich Behavior Recognition Based on Deep Learning

Yusheng Duan
School of Electric Information
and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: Ostrich behavior is a key sign of their development and health. Quickly and accurately identifying it is crucial for growth monitoring and disease prevention. Computer vision technology, being real-time and non-contact, is widely used in livestock behavior recognition. But current methods, mainly for common livestock in simple settings, have drawbacks.This paper presents a method for ostrich behavior recognition using YOLOv7-MG. It aims to boost recognition efficiency and precision. Images of ostrich behavior are gathered from actual farms to create a dataset. The MobileOne network replaces the backbone of YOLOv7 to cut down computation and model size. Also, a GAM module is added to improve feature extraction in complex situations.The proposed method does better than YOLOv7 and other cattle behavior recognition systems. It has a relatively small model memory footprint and can precisely identify ostrich behavior. This lays the groundwork for ostrich disease prevention and management.

Keywords: Image recognition; Deep learning; Ostrich Behavior Recognition; image processing; Intelligent Farming

1. INTRODUCTION

Nowadays, intelligent breeding of livestock and poultry has increasingly become a prominent research area. Ostriches, in particular, possess high economic value, not only due to their valuable products such as meat, eggs, and feathers which are in significant market demand [1], but also because of their rapid growth rate, relatively lower costs, and shorter breeding cycles. However, ostrich farming is not without its challenges. Disease prevention and control is a crucial aspect, as ineffective management can result in substantial financial losses. Traditional approaches like manual inspection and symptomatic preventive medicine have proven to be insufficient, being hampered by low efficacy, high costs, and a narrow monitoring range .

The emergence of artificial intelligence and computer vision technologies has brought new possibilities to the livestock and poultry breeding industry. In the domain of behavior recognition, several techniques have been explored. For instance, Nasirahmadi and colleagues employed deep learning techniques such as Faster R-CNN to identify the postures of pigs [2]. Liu et al. utilized an enhanced YOLO v3, optimizing it with the amplitude iterative pruning approach to achieve a 79.9% recognition accuracy for cow feeding behavior [3]. Kim et al. adopted the YOLOv3 and YOLOv4 models to identify the eating behavior of suckling pigs [4]. Wang et al. proposed an enhanced YOLOv5-based cow behavior detection system for accurately identifying cow behavior during estrus [5].

Nonetheless, these existing methods are mainly designed for relatively simple and less complex environments, and they often lack the adaptability required for the more intricate and dynamic conditions of ostrich farming. The free range of activity for ostriches is considerably larger compared to that of pigs and cows, with a greater likelihood of other objects being present in their activity area, thus making the living environment more complex.

Therefore, this paper centers around ostrich behavior recognition. Initially, data is meticulously gathered from the actual ostrich breeding environment to construct a comprehensive dataset. Subsequently, taking YOLOv7 as the foundation, the model is refined by substituting the YOLOv7 backbone network with the lightweight MobileOne backbone network, effectively reducing the number of parameters and computational load . Finally, the Global Attention Mechanism (GAM) module is incorporated into the header to augment the model's capacity to extract features from complex surroundings. The ultimate objective is to propose a lightweight and efficient ostrich behavior recognition model that can serve as a valuable reference for the intelligent management of ostrich farming, enhancing both productivity and the overall health and well-being of the ostrich population.

2. MATERIAL AND METHODS

2.1 Experimental Data Acquisition and Processing

The experimental dataset was collected from an ostrich farm. The area where the ostriches were naturally active was chosen as the data collection site to ensure the universality of the data. Mobile phones were then utilized to take pictures of the ostriches from various angles at a suitable distance outside the farm.

The video recorded contains common postures of ostriches. A third-party Python module was employed to extract one frame every ten frames and export the movie as JPG files. To avoid the overfitting problem caused by the high visual similarity of adjacent frames and the single ostrich feature, data cleaning was carried out. After manual inspection and screening, 800 photos were retained for further processing and experimentation. Subsequently, some images with relatively low quality and excessive similarity were removed.

2.2 Data Enhancement

In order to strengthen the robustness of the model, 800 negative samples were included in the ratio of positive and negative samples 1:1, totaling 1600 pictures as the original image dataset. It is quite likely to result in issues like an unstable training process and overfitting of the model if these 1600 photos are utilized straight for training and testing. This

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

work expanded the original dataset from 1600 to 6400 pictures using data augmentation techniques, such as varying image brightness and introducing Gaussian noise, to improve model performance and generalization capacity.

Figure. 2 Improved YOLOv5s

2.3 Algorithmic Modeling and its Improvement

Chien-Yao Wang and Alexey Bochkovskiy et al. created the YOLOv7 model in 2022. The model incorporates a number of techniques, such as model reparameterization, mod-el scaling based on tandem models [6], and E-ELAN (Extended Efficient Layer Aggregation Network) [7]. The backbone network, header network, and prediction network are the three primary parts of the YOLOv7 network. The MP module, which does preliminary feature extraction on the input pictures, the E-ELAN module, and the CBS module make up the majority of the backbone network.

The primary components of the head network are the MP, ELAN-H, Concat, and Spatial Pyramid Pooling and Convolutional Spatial Pyramid Pooling Sppcspc modules. These modules combine and improve the incoming characteristics of the backbone net-work to extract the most important information. After using the Rep structure to help train the prediction network, the number of image channels for the head network's output fea-tures is adjusted using 1x1 convolution, and the predicted correlation data is eventually acquired.The network structure of the YOLOv7-MG lightweight ostrich behavior recognition model proposed in this paper is shown in Figure. 2.

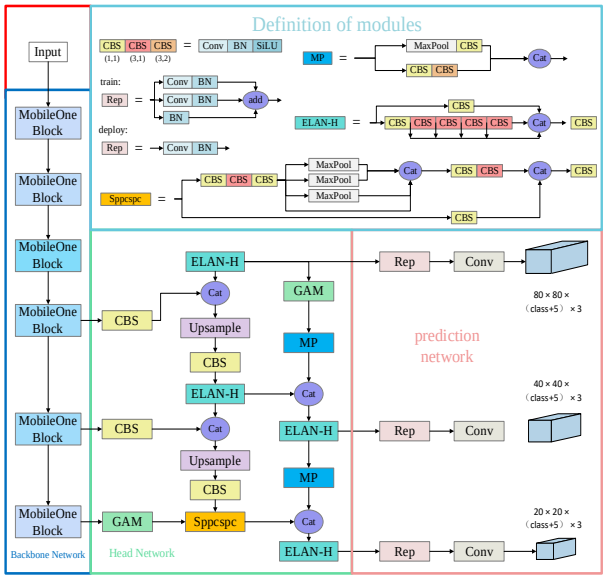

Figure. 2 YOLOv7-MG structure diagram

2.3.1 *Replacement of the backbone network*

The applicability and practical usefulness of the model application to the mobile terminal may be shown in the recognition of ostrich behavior and further study in this ar-ea. As a result, the model must have less complexity; in other words, it must be light-weight. In this study, we replace the original YOLOv7 backbone network with a light-weight one called MobileOne[8]. The original goal of MobileOne, a lightweight convolu-tional neural network, was to maximize model performance and compute economy in situations involving mobile or edge applications. Several MobileOne Blocks make up the MobileOne paradigm, and these MobileOne Blocks are interconnected. The step size and the number of output channels are the sole distinctions between the various MobileOne Blocks that make up the MobileOne paradigm. Depthwise Convolution + Pointwise Con-volution provides the fundamental structure of the MobileOne Block structure, which in-corporates the parameterization concept from RepVGG [9]. Fig. 2 depicts the MobileOne Block construction.

During training, the depth convolution section consists of k blocks of 3×3 depthwise convolution, one branch of 1×1 depthwise convolution, and one batch normalization (BN) branch in parallel. The point convolution section includes k blocks of 1×1 pointwise con-volution and one BN branch in parallel. In contrast, during inference, the MobileOne Block structure has no branches and follows a streamlined architecture. This design ensures ef-fective feature extraction during training while providing fast inference and low model memory usage during prediction.

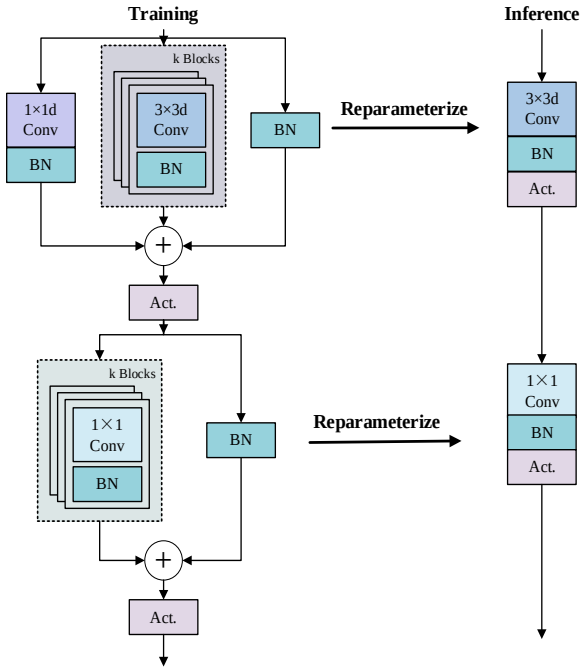

Figure. 3 MobileOne Block structure diagram

2.3.2 *Introduction of attention mechanisms*

It can be observed from the dataset photographs that ostriches provide shade to each other, as well as to themselves, due to the surrounding fences and trees. As a result, the complexity of the environment presents a greater challenge for the model's ability to ex-tract features. The attention mechanism helps the network model focus more on the im-portant aspects

of the image. The popular SENet[10] and CBAM [11] attention modules (Convolutional Block Attention Module

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

and Squeeze and Excitation Network, respectively) The connections between dimensions are weakened when Attention Modules overlook the interactions between space and channels. On the other hand, by improving the interaction of information in global dimensions, GAM [12] can lessen the dispersion of important in-formation in features and enhance the network's overall capacity for critical feature ex-traction. It is composed of a spatial attention sub-module (e.g., Fig. 6) and a channel atten-tion sub-module (e.g., Fig. 5), as seen in Fig. 4.

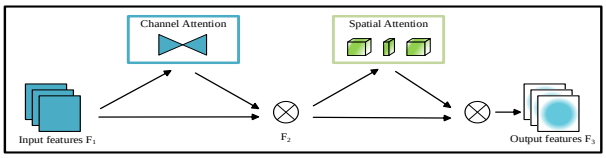

Figure. 4 GAM structure diagram

In Figure 4, the results of the GAM attention module, denoted as F_3, are obtained by multiplying the input feature map F_1 by the channel attention map $M_c(F_1)$ element-wise, followed by multiplication of the intermediate feature map F_2 with the spatial attention map $M_s(F_2)$ element-wise. The diagram clearly indicates these element-wise multiplication operations, highlighting the importance of each attention mechanism.

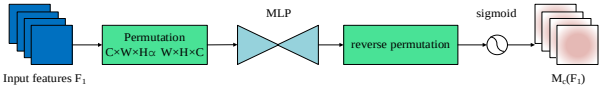

Figure. 5 Structure of the channel attention sub-module of the GAM module

As shown in Figure 5, the channel attention submodule first retains the information across three dimensions (channel, spatial width, and spatial height) by using 3D permu-tation. Then, it magnifies the cross-dimensional channel-spatial dependencies with a multi-layer perceptron (MLP). The MLP employs an encoder-decoder structure and re-duces the dimensionality through a reduction ratio to improve computational efficiency and performance.

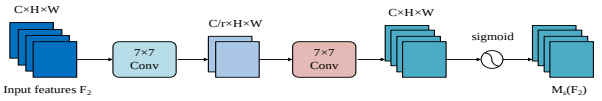

Figure. 6 Structure of the spatial attention sub-module of the GAM module

As shown in Figure 6, the channel attention submodule in the GAM module consists primarily of a 3D transformation operation, which allows it to preserve the spatial, chan-nel, and depth-related information of the input features. This transformation is followed by a two-layer Multi-Layer Perceptron (MLP), which enhances the dependencies between the channel and spatial dimensions. The spatial attention submodule, on the other hand, is composed of two 7x7 convolutional layers designed to efficiently fuse spatial infor-mation. In the first convolutional layer, the number of channels is reduced from C to C/r, where r is a reduction ratio hyperparameter that controls the degree of dimensionality re-duction to balance between computational efficiency and information retention. Together, these operations in the GAM module enable the network to focus on regions with signifi-cant contextual importance, while reducing the dispersion of meaningful feature infor-mation, thus improving the overall performance of the model.

3. EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental Environment and Parameter Settings

The operating environment for this experiment features a 12th Gen Intel(R) Core(TM) i5-12400F CPU, 16GB of RAM, and an NVIDIA GeForce RTX 3060 GPU. The experiments are conducted within the PyTorch deep learning framework, with a CUDA version of 11.1.

During the training, the initial learning rate (Base_lr) is configured as 0.01, the weight decay is set to 0.0005, the optimizer (Optimizer) is chosen as SGD, and the batch size (Batchsize) is set to 8.

3.2 Experimental Results

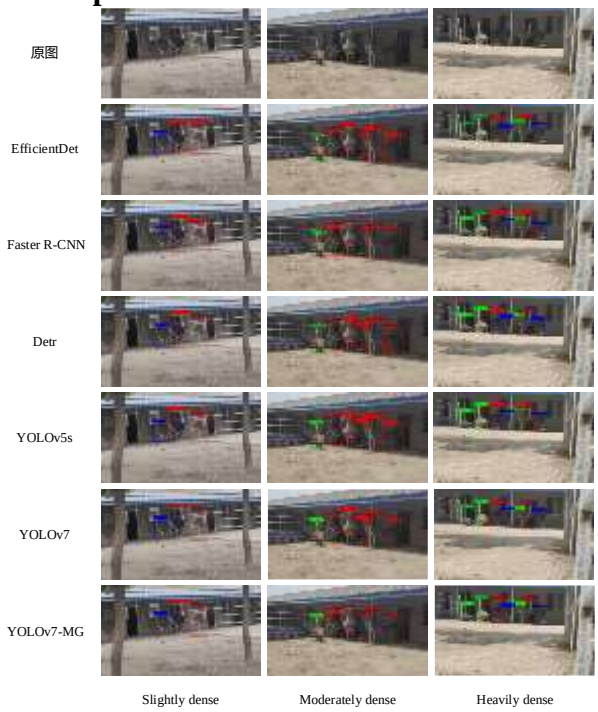

Figure. 7 Comparison of ostrich behavioral recognition results in environments with different densities

Figure. 7 shows a comparison of the results of ostrich behavior recognition in three different densely populated environments (slightly dense, moderately dense and heavily dense). Under the slightly dense environment, all the ostrich's behaviors were accurately identified. In the moderately dense environment, most of the ostrich behaviors were accurately identified, and the accuracy of individual ostrich behaviors that were not too badly occluded decreased slightly. In the heavily dense environment, two ostriches' behaviors were missed due to severe occlusion, and the rest of the ostrich behaviors were accurately identified. Each model was able to identify the target that should be identified, and there were differences in the level of confidence, accuracy of target frame localization, and detection speed between the different models. Although EfficientDet, FasterR-CNN, and DETR were slightly less accurate in localization, the overall performance was still quite good.YOLOv5s and YOLOv7 were balanced and efficient in terms of speed and accuracy.YOLOv7-MG was the best performer, with optimal localization and recognition accuracy, and was the fastest in terms of recognition speed. This indicates that the robustness

of the model in this study is high, with only a few misrecognitions in heavily dense environments.

4. CONCLUSION

In this study, an ostrich behavior recognition approach based on enhanced YOLOv7 is proposed to identify the daily behaviors of farmed ostriches. To enhance the model's efficiency and adaptability, the YOLOv7 backbone network is replaced with the lightweight MobileOne backbone network, reducing computational complexity and parameter numbers. Subsequently, a global attention mechanism (GAM) module is incorporated into the head network to improve the model's feature extraction capabilities in complex environments.

Compared to the original YOLOv7 model, the proposed method demonstrates notable improvements in various aspects. It achieves a significant enhancement in mean average accuracy and recognition speed, while also optimizing the model size. The experiments are conducted using a self-built dataset, as there is currently no relevant public dataset available.

When compared to other popular models such as EfficientDet, Faster R-CNN, Detr, YOLO5s, and the original YOLOv7, the approach presented in this research shows superiority in both identification speed and mean average accuracy. This indicates that the YOLOv7-MG model proposed in this paper outperforms existing models in recognizing ostrich behavior in complex environments, with better robustness.

However, it should be emphasized that the acquisition of high-quality datasets of ostrich behavior images remains a major challenge in the development of ostrich smart farming. Future efforts should focus on collecting a sufficient number of high-quality images of abnormal ostrich behaviors, which will lay the foundation for the advancement of ostrich abnormal behavior detection and the intelligent prevention and control of ostrich diseases.

5. REFERENCES

[1] Medina F X, Aguilar Moreno E. Ostrich meat: nutritional, breeding, and consumption aspects. The Case of Spain[J]. 2014.

[2] Nasirahmadi A, Sturm B, Edwards S, et al. Deep learning and machine vision approaches for posture detection of individual pigs[J]. Sensors, 2019, 19(17): 3738.

[3] Liu Yuefeng, Bian Haodong, He Yingjie, et al. A Recognition Method for Multi-target Cow Feeding Behavior Based on Amplitude Iterative Pruning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53 (02): 274 - 281.

[4] Kim M J, Choi Y H, Lee J, et al. A deep learning-based approach for feeding behavior recognition of weanling pigs[J]. Journal of animal science and technology, 2021, 63(6): 1453.

[5] Wang R, Gao Z, Li Q, et al. Detection method of cow estrus behavior in natural scenes based on improved YOLOv5[J]. Agriculture, 2022, 12(9): 1339.

[6] Gao P, Lu J, Li H, et al. Container: Context aggregation network[J]. arXiv preprint arXiv:2106.01401, 2021.

[7] Dollár P, Singh M, Girshick R. Fast and accurate model scaling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 924-932.

[8] Anasosalu Vasu P K, Gabriel J, Zhu J, et al. An Improved One millisecond Mobile Backbone[J]. arXiv e-prints, 2022: arXiv: 2206.04040.

[9] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[11] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[12] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv:2112.05561, 2021.