Research on Sheep Recognition Algorithm Based on Deep Learning in Animal Husbandry

Chao Ma, Xiaolin Sun, Chunxia Yao, Minglu Tian, Linyi Li *
Shanghai Academy of Agricultural Sciences, Shanghai, Shanghai, China
*Corresponding author’s e-mail: lly@saas.sh.cn

Abstract. At this stage, the domestic animal husbandry industry has a very important support role in the domestic market economy, and the sheep industry has received extensive attention as one of the important industries. Limb movements and activities can directly reflect the adaptability of sheep to the breeding environment and conditions, and provide a richer scientific and technical experience for sheep breeding. The sheep target detection is an important prerequisite for grasping the movement behavior of sheep and providing a richer scientific and technical experience for sheep breeding. Therefore, how to efficiently and accurately identify and detect sheep has become the key to the development of sheep industry at this stage. In this paper, the Faster-RCNN neural network model based on the Soft-NMS algorithm is studied, which realizes the real-time detection and positioning of sheep under complex breeding conditions, and improves the accuracy of detection while ensuring the detection speed. Experiments show that the proposed detection model can detect sheep with 95.32% accuracy and mark the location of the target in real time, which provides an effective data foundation for sheep behavior research and helps promote the development of high-tech animal husbandry effect.

1. Introduction
At present, the development of China's animal husbandry is in a fierce state. In the modern and scientific development of sheep farming, it is very important to acquire and process the specific information of sheep targets. In order to objectively judge the breeding status of sheep animals, it is necessary to accurately obtain the activity information and physical health status of sheep individuals in real time. Through reasonable intervention of artificial forces, the resources can be maximized, and a stable guarantee can be provided for the sustainable development of the sheep industry. Therefore, in order to improve the efficiency and accuracy of the sheep farming industry and ensure the breeding effect and production quality, it is necessary to identify and detect the sheep targets in the complex breeding environment.

Now, China's research in the field of deep learning is gradually increasing. Target detection based on deep learning has been widely used, such as road vehicle detection, face recognition, defect detection, etc. At the same time, domestic researchers have gradually applied deep learning to animal husbandry, conducted efficient detection of live animals, and harvested more obvious research results, which have been widely used. In 2013, Simona et al. make full use of the advantages of computer vision and adopt the Viola Jones algorithm to realize the lying posture recognition of cattle targets [1]. Based on the gait information of cattle, Blackie et al. use effective analysis techniques to detect diseases such as foot ulcers or ulcers in cattle [2]. In 2014, Viazzi et al. use vision technology to achieve efficient image segmentation and successfully identify the back pose of cattle targets [3]. In 2015, Thorup et al. realize the judgment of the posture of locomotion by extracting the features of...
animal target's ambiguity [4]. In 2016, He Dongjian et al. propose an improved target detection method, and build a model for the target loop search environment to achieve calf detection under complex breeding conditions [5]. In 2017, Liu Yanqiu et al. design a wearable monitoring device for the prenatal behavior of ewes, which can achieve accurate identification and efficient classification of prenatal behavior, and the detection accuracy rate is as high as 82.69% [6]. In 2018, Xue Yueju et al. make algorithm improvements based on Faster R-CNN, adding residual structure and center loss supervision signal, which greatly improves the detection accuracy of the entire model and has real-time characteristics [7].

Based on the excellent performance of deep learning in the field of target detection, this paper proposes a Faster-RCNN neural network model based on the Soft-NMS algorithm for the challenges and practical application needs in animal husbandry. Using the excellent detection ability of deep learning in complex scenes, the detection of sheep targets under complex breeding conditions is achieved, and the specific location of sheep is marked. Through the establishment of image data sets of sheep, the training and testing results show that this method can greatly improve the detection and recognition effect, which reflects the scientific concept of animal husbandry on the road of intelligent development, and at the same time creates economic benefits for breeding.

2. Faster R-CNN

In this paper, VGG16Net is used as the feature extraction network of Faster-RCNN. The structure of Faster R-CNN is shown in figure 1, which mainly includes four parts: convolutional layer, RPN, ROI pooling, classification and regression. The size of the original image is $X \times Y$, and the size input into the Faster-RCNN model is converted to $W \times H$. The image first passes through 13 conv layers, 13 relu layers and 4 pooling layers. In order to ensure that the size of the input and output matrices will not be affected by the convolutional layer in the feature extraction network structure, we set the size of all convolutional layers here to $3 \times 3$, and the size of all pooling layers to $2 \times 2$. After the image passes through a pooling layer, the size will change from $W \times H$ to $(W/2) \times (H/2)$. There are four pooling layers in the entire feature extraction, so the final output feature map size is $(W/16) \times (H/16)$.

![Figure 1. Schematic diagram of Faster R-CNN based on VGG16Net.](image)

2.1. RPN

In the Faster R-CNN network model based on VGG16Net, the input image is extracted into the RPN after feature extraction, and the foreground anchors are used for regression of the candidate frame. At the same time, a more accurate candidate box is calculated with a certain amount of translation and scaling, and then enters ROI pooling to collect and summarize the candidate boxes. In ROI pooling, a fixed-size target recognition area is generated by normalization, after that it is sent to classification. Finally, the class to which the candidate box belongs is obtained through the fully connected layer and
softmax. In addition, through the bounding box regression, the degree of deviation of the candidate frame from the real position is obtained, and this is used as the regression basis of the target detection frame, so that the entire target detection is more accurate.

Among them, the bounding box regression mainly uses some means to adjust the foreground anchors (FA) to make it more consistent with the ground truth (GT). Figure 2 shows the schematic diagram of window, including the window GT of the target sheep and the window FA extracted by the model, but there is a certain error in the position of the window FA in the actual detection process, which causes all the information in the detection window to not completely represent the sheep identified. Therefore, we add window N and improve the matching degree of window GT and window FA through the following transformation.

The four-dimensional vectors \((x, y, w, h)\) used represent the center coordinates, width, and height of the window, which can be obtained by translation:

\[
N_x = FA_x \cdot d_x (FA) + FA_x
\]

\[
N_y = FA_y \cdot d_y (FA) + FA_y
\]

Then through a certain zoom can be obtained:

\[
N_w = FA_w \cdot \exp (d_w (FA))
\]

\[
N_h = FA_h \cdot \exp (d_h (FA))
\]

After the above pan and zoom, the window closest to GT can be finally obtained:

\[
(GT_x, GT_y, GT_w, GT_h) \approx (N_x, N_y, N_w, N_h)
\]

2.2. Soft-NMS

The Soft-NMS algorithm implements frame removal in a relatively soft way. Instead of clearing the detection score of the adjacent detection frame that overlaps with the detection frame \(M\), an attenuation operation is used. In addition, the complexity of the Soft-NMS algorithm is not high, which can reduce the pressure of the original model training.

In the traditional NMS algorithm, a fixed recording function is usually used to describe the basic process of the algorithm, and its expression is:

\[
S_j = \begin{cases} 
S_j, & \text{iou} (M, b_j) < N_i \\
0, & \text{iou} (M, b_j) \geq N_i 
\end{cases}
\]
The NMS algorithm judges the adjacent detection frame with a clear threshold to determine the removal of the detection frame. The Soft-NMS algorithm optimizes the recording function:

\[
S_j = \begin{cases} 
S_{j,iou(M, b_j) < N_i} \\
S_{j, (1-iou(M, b_j)), iou(M, b_j) \geq N_i}
\end{cases}
\]  

Among them, \(N_i\) is the overlap threshold between the detection frames. When \(iou(M, b_j) \geq N_i\), that is, the degree of overlap between the adjacent detection frame and the detection frame M exceeds the threshold value \(N_i\), the detection score of the detection frame at this time shows a trend of attenuation. At the same time, the attenuation of the detection frame closer to M is much greater than the detection frame farther away. Since the score reset function in the expression is segmented on the condition of \(N_i\), when it is greater than \(N_i\), it will cause a large fluctuation in the final generated sequence, which will affect the accuracy of the result.

In order to solve this problem, a continuous score reset function is proposed, which can make the correct attenuation effect according to the overlapping degree of the detection frame. The improved score reset function expression is:

\[
S_j = S_j e^{-\frac{iou(M, b_j)^2}{\sigma}}
\]

3. Sheep target recognition detection based on deep learning

3.1. Data set creation
In order to verify the accuracy of the sheep target detection model proposed in this paper, we first establish a data set for training and testing. All the experimental data used in this paper are from ordinary sheep in the pasture. In order to ensure the scientific nature of the data, we use multiple angles for video shooting, and collect a total of 20 pieces of daytime herd activity videos. Through certain technical means, all the sheep activity videos collected are framed to obtain sheep images. There are a large number of invalid images in the data, which need to be rigorously screened out to get rid of them, and finally get 3480 daytime images. All the images obtained are marked in detail, mainly including the position of the sheep and the number of sheep targets, and XML files with the same name as the images are generated. In the experiments in this paper, the collected image data is randomly generated in the ratio of 8:2 to the corresponding training set and testing set, which are used for network model training and accuracy testing, respectively.

3.2. Experimental platform construction and parameter setting
In this paper, we study sheep target detection based on deep learning. Before training and testing, we need to configure the experimental environment accordingly to meet the detection requirements of the experiment from both hardware and software aspects. The experimental system configuration is as follows: the hardware device model is HP Z8 G4 (Z3Z16AV-SC001), in which the CPU type is Intel Xeon Bronze 3104, the memory size is 16GB, the operating system is Ubuntu 16.04 64bit, the deep learning framework is Caffe, and the programming language is Python. In training, we set the coverage domain threshold to 0.5, the non-maximum suppression threshold to 0.7, and the initial learning rate to 0.001. After 1000 trainings, an effective network model is obtained.

3.3. Evaluation index
In order to accurately evaluate the performance of the model, we use precision (P), recall (R), and average accuracy AP as evaluation indicators. For a given test set, TP represents the predicted number of targets, which is actually the number of targets. FP represents the predicted number of targets, and actual non-targets. TN represents the predicted number of non-targets, and the actual number of non-
targets. FN means predicting non-targets, which is actually the number of results of targets. P and R can be expressed as:

\[ P = \frac{TP}{TP + FP} \]  
\[ R = \frac{TP}{TP + FN} \]

AP refers to the average value of P under different values of R. The calculation formula is:

\[ AP = \frac{1}{k} \sum_{j=1}^{k} P(R_j) \]

3.4. Experimental results
In order to comprehensively evaluate the detection performance of the algorithm Faster R-CNN used in this paper, the traditional NMS and Soft-NMS are iteratively trained and tested by alternating optimization training methods using the data set established in this paper. The visualization of the detection results is shown in figure 3, and the comparison results of the experimental AP are shown in table 1. The Faster R-CNN network model based on the Soft-NMS algorithm has obvious effects on the visualization results of sheep target detection. Compared with the NMS algorithm, the accuracy of target detection is higher, which shows that Faster R-CNN based on Soft-NMS algorithm is more suitable for the detection of sheep targets.

| Algorithm                        | AP     |
|----------------------------------|--------|
| Faster R-CNN (NMS)               | 94.13% |
| Faster R-CNN (Soft-NMS)          | 95.32% |

From the comparison table of detection results above, it can be seen intuitively that the detection accuracy of sheep targets based on Soft-NMS algorithm is 1.19% higher than that based on NMS algorithm. Therefore, the Faster-RCNN model based on Soft-NMS has a better effect in detecting sheep targets.

4. Conclusion
The Faster-RCNN model based on Soft-NMS has good scalability. In this paper, the sheep target data set is made, and the sheep target detection is achieved through effective training and testing. The test results show that it is feasible to build a Faster R-CNN network model using Soft-NMS algorithm to detect sheep targets, which can effectively detect the specific location of sheep. The detection system based on the model designed in this paper has a very broad research prospect in the sheep breeding industry. The development of a complete detection system can be achieved by building an image
acquisition platform and designing a host computer, which has an important role in the development of
the entire livestock industry in terms of intelligence and technology.

Acknowledgments
This work was supported by Shanghai Agriculture Applied Technology Development Program, China:
Fengxian District Malu Village Village Revitalisation technology leads demonstration village
construction (Topic 5) (Shanghai Agricultural Science Promotion (2019) No. 2-3-3).

References
[1] Porto S. M. C., Arcidiacono C., Anguzza U., et al. (2013) A Computer Vision-Based System for
The Automatic Detection of Lying Behaviour of Dairy Cows in Free-Stall Barns. J. Biosystems Engineering, 2: 184-194.
[2] Blackie N., Bleach E. C. L., Amory J. R., et al. (2013) Associations between Locomotion Score
and Kinematic Measures in Dairy Cows with Varying Hoof Lesion Types. J. Journal of Dairy Science, 6: 3564-3572.
[3] Viazzi S., Bahr C., Van Hertem T., et al. (2014) Comparison of a Three-Dimensional and Two-
Dimensional Camera System for Automated Measurement of Back Posture in Dairy Cows. J. Computers and Electronics in Agriculture, 100: 139-147.
[4] Thorup V. M., Munksgaard L., Robert P. E., et al. (2015) Lameness Detection Via Leg-
Mounted Accelerometers on Dairy Cows on Four Commercial Farms. J. Animal, 10: 1704-1712.
[5] He D. J., Meng F. C., Zhao K. X., et al. (2016) Recognition of Calf Basic Behavior Based on
Video Analysis. J. Journal of Agricultural Machinery, 9: 294-300.
[6] Liu Y. Q., Wu P., Zhang L. N., et al. (2017) Feature Analysis and Recognition of Ewe’s
Behavior Prior to Delivery—Based on a Wearable Detection Device Architecture. J. Agricultural Mechanization Research, 9: 163-168.
[7] Xue Y. J., Zhu X. M., Zheng C., et al. (2018) Lactating Sow Postures Recognition from Depth
Image of Videos Based on Improved Faster R-CNN. J. Transactions of the Chinese Society
of Agricultural Engineering, 9: 189-196.