A REAL-TIME SHEEP COUNTING DETECTION SYSTEM BASED ON MACHINE LEARNING
/
一种基于机器学习的羊群实时计数检测系统

Xuefeng DENG *1), Song ZHANG 2), Yi SHAO 1), Xiaoli YAN 3) [1]
1) College of Information Science and Engineering, Shanxi Agricultural University, Taigu / China
2) Agricultural Engineering and Information technology, Inner Mongolia Agricultural University, Hohhot / China
3) Taiyuan Jinshan Middle School, Taiyuan / China
Tel: 13546632115; E-mail: dxf@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-67-08

Keywords: *YOLOv5, Target detection, Counting sheep, Machine learning*

ABSTRACT
With the development of modern breeding industry, it is very important to count sheep accurately. In the past, herdsmen used manual statistics to count and manage sheep, which was time-consuming, laborious and often had large errors. In recent years, machine learning methods are widely used in automatic target recognition, which can replace manual labor. This system is based on YOLOv5 algorithm for sheep counting management. The counting of sheep was controlled by two - way counting. This improves the accuracy of counting, saves a lot of manpower and material resources for herdsmen, and greatly promotes the management of animal husbandry.

摘要
在大面积的养殖牧场，随着现代养殖业的不断扩大，如何对羊群进行精确计数显得尤为重要。在过去，牧民们用人工统计的方法对羊群进行计数管理，该方法既费时又费力，还经常存在较大的误差。近年来，机器学习的方法大量用于目标自动识别，可以代替人工劳动。本系统基于YOLOv5算法对羊群进行计数管理，通过双向撞线计数法，采用两条线控制羊的计数，提高了计数精确度，为牧民节省了大量的人力、物力资源，极大地推动了养殖牧业的管理。

INTRODUCTION
The accurate counting of sheep plays a decisive role in intelligent management of sheep. With the rapid development of China's economy, the improvement of people's living standards leads to the prosperity of animal husbandry, and the management of sheep in the pasture becomes tricky. The introduction of intelligent management system can effectively solve the past herdsmen using artificial statistical methods exist shortcomings. Manual management mode not only wastes a lot of human resources, but also its statistical effect is often unsatisfactory. As a basic management module, sheep counting statistics is a core content of intelligent management system.

The counting problem is part of the statistical problem. In the era of big data, statistical problems have been exposed in many fields. For example, the statistical management of dense population (*Lu et al., 2019*), the statistical management of traffic flow (*Jiang et al., 2012*), the statistical management of biological cells (*Chen et al., 2019*), the group characteristics analysis (*Anlan et al., 2019*) and the statistical management of library reader flow (*Guo et al., 2019*) and so on, almost covering all walks of life. The corresponding statistical algorithms are also varied and popular in all kinds of places. Among the traditional statistical management techniques, there is a statistical method based on infrared sensor, which has good detection performance when there is a certain distance between the target objects. However, in the case of occlusion between target objects, the detection accuracy will be reduced and it is not suitable for group class counting. The statistical method based on background features is suitable for detecting and counting when the feature color channel extracted from the target object differs greatly from the background color. Therefore, the statistical management effect is poor when the two colors are similar. Based on the statistical method of pressure sensing, it can only estimate the number of objects detected. And the laying of equipment needs to be carried out in the fixed site, which has the disadvantages of low detection accuracy, poor scalability and high physical cost.

---
[1] *Xuefeng Deng*, assoc.prof, PhD; Song Zhang, M. Stud.; Yi Shao, undergraduate; Xiaoli Yan, M.Eng.Lec., graduate.*

With the continuous progress of machine learning, artificial intelligence and other technologies. By means of machine vision and target detection, effective counting statistics can be carried out. Deformable Parts Model (DPM), Region based Convolutional Neural Network (R-CNN) (*Sun et al., 2018*), Fast Region based Convolutional Neural Network (Fast R-CNN) (*Girshick et al., 2015*), Faster Region based Convolutional Neural Network (Faster R-CNN) (*Chen et al., 2018*), Single Shot MultiBox Detector (SSD) (*Song et al., 2006*), You Only Look Once (YOLO) (*Deng et al., 2021*) and other technologies have been produced in real-time target detection. DPM technology makes use of artificially designed templates to traverse the target detection graph in turn, but its running speed is slow and its robustness is poor. R-CNN firstly extracts multiple candidate boxes from the target image and then processes them. Its speed is slow, but the prediction result is relatively accurate. In addition, the improved Fast R-CNN and Faster R-CNN technologies both belong to two-stage models, and the main problem exposed is the slow speed and easy loss of image information. YOLO technology belongs to the end-to-end single-stage model, which is in one step compared with the two-stage model and greatly improves the processing speed. In recent years, with the continuous maturity of YOLO technology, the current YOLOv5 model improves the defect of YOLO's low recognition performance on small objects and small groups. It can almost achieve the real-time effect of naked eye observation, and has good robustness in real-time counting.

In this experiment, YOLOv5 technology was used to count sheep in pasture, and a real-time detection system was built to count sheep. This reduces the workload of herders' management and improves the accuracy of sheep counting. The system has been tested and proved to be accurate, robust, real-time and expandable. This method can be extended to the sheep management system, as the core module of the system to detect the number of sheep, and can also be used in the pasture management system.

MATERIALS AND METHODS

Sheep data set collection

The Inner Mongolia Autonomous Region has a grassland area of 86.667 million hectares, of which 68 million hectares are effective natural pastures, accounting for 27 percent of the national grassland area. Therefore, the animal husbandry of Inner Mongolia Autonomous Region is more developed and there are many herdsmen. Cattle and sheep breeding has become the pillar industry of the region. In this experiment, 28 small-tailed han sheep were taken as test objects and 29 videos were collected, among which 28 videos corresponded to the data results of each sheep, and the last one was the verification set. In practice, there are often multiple sheep in the video because the sheep are difficult to control. Therefore, it is necessary to clip the video of the data set first, delete the video clips in the case of no sheep and unclear sheep, and then save the remaining video clips by frame hopping. When the frame number is too small, the posture of the target to be detected may be similar, which leads to data set duplication. However, when the frame number is too large, the sample data will be less, so that the required accuracy of the experiment can not be achieved. In this paper, the frequency of extracting a picture every 15 frames is saved as the test data set. Because this design involves target tracking, it is necessary to quickly capture sheep in the first few frames of the image, and at the same time, it is necessary to add various poses of sheep into the training set. The specific situation is divided into four situations: sheep just appeared in the channel (sheep head), sheep front, sheep side, and sheep covered, as shown in Figure 1.

Introduction to YOLO algorithm

YOLO technology, proposed by Redmon J., solves the detection of target objects as a regression problem. It takes the whole picture as the input part to get the location and category information of the target object, which belongs to the single-stage model. Compared with the traditional two-stage model, it has the characteristics of faster detection speed, lower background misjudgment rate and good extensibility. Specifically, it can still have a high degree of detection rate in abstract works of art. However, when encountering herds, that is, a detection grid contains multiple objects to be detected, the principle based on YOLO can only detect one of the objects to be detected. So YOLO is less accurate for small groups of objects (*Redmon et al., 2016*).

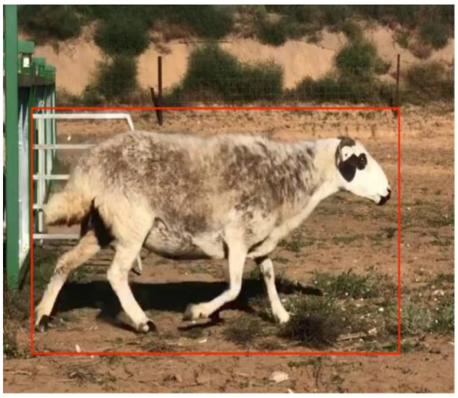
On the basis of YOLO, YOLOv2 technology adopts joint training algorithm to expand the detection target set to 90,000 kinds, which can identify more target objects (*Redmon et al., 2017*). The batch normalization method is used to solve the problems of gradient disappearance and gradient explosion in the process of inverse feedback. The high resolution image classifier was adopted to improve the effect of YOLO on the sudden resolution switch in the process of training CNN convolutional layer.

At the same time, it also draws on the advantages of Faster R-CNN technology and uses the prior frame to predict and locate the target object. The mesh parameters are dynamically changed according to the number of iterations in the training set to improve the accuracy and detection speed of the system. It supports the input of images of various sizes and classifies the object with decision tree. Compared with YOLO, it is more accurate, faster and more robust. YOLOv3 technology adopts darknet-53 network structure, and based on YOLOv2, FPN is introduced to carry out multi-scale detection of target detection objects, so as to enhance the detection rate of small objects (*Redmon et al., 2018*). The problem of gradient explosion of neural network is further improved by using residual network. In the training of loss function, dichotomous cross entropy is used to replace the original sum of squares of error and reduce the training error. This makes YOLOv3 can not reduce the original recognition speed, in the detection rate of small objects and close objects have a large level of improvement. The YOLOv4 model is the result of technical integration (*Bochkovskiy et al., 2020*). Almost every part of YOLOv3 model is optimized and improved, and an efficient and powerful target detection model is proposed. The technology of continuous stacking and parameter adjustment makes its detection faster and more accurate.

YOLOv5 model combines machine vision technology, which greatly improves the speed and model size compared with the previous 4 versions (*Jia et al., 2021*). The enhancement of training data such as Mosaic effectively solves the low efficiency of small target detection in the original model. The anchor frame in YOLOv5 adopts the form of automatic learning of training data, so that the code can be automatically run without the waste of resources of manual modification. Cross Stage Paritial Network (CSPNet) is used to alleviate the problem of gradient disappearance and improve the reasoning speed and accuracy of the model. At the same time, the model size has been greatly reduced to a lightweight model size. YOLOv5 was almost 4 times faster than YOLOv4 in model training. In addition, its model size is small, which is conducive to the rapid construction of the model and more suitable for the real-time target detection environment. The overall performance of YOLOv5 is somewhat improved over all previous versions, with greater speed and flexibility. So that the target detection technology can be more widely used in large-scale production development.

(a) Sheep head posture

(b) Sheep frontal posture



(c) Sheep side posture

(d) Sheltered sheep posture

Fig. 1 - Sheep posture capture picture

YOLOv5 detection process

YOLO algorithm is One of the most widely used one-stage target detection algorithms. As the latest technology in the YOLO series, YOLOv5 has the characteristics of higher recognition accuracy, speed and light weight. It is currently the most advanced version of the YOLO series.

YOLOv5 is the same as YOLOv4 in structure. It consists of the input layer, BackBone layer, Neck layer, and output layer. The input layer preprocesses the input image, typically resetting the image to 608×608 or 416×416.

In YOLOv4, images need to be filled with black edges during reshape. However, in YOLOv5, the adaptive filling method is used, in order to minimize the use of black edges for filling to reduce the amount of computation. The image filling process is shown in Figure 2.

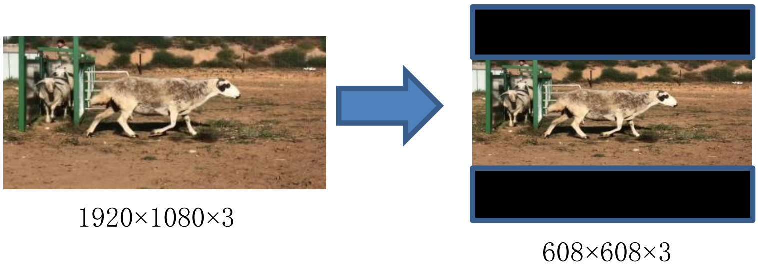


Fig. 2 - Image filling process

The filled pictures are sent to backbone network for preliminary feature extraction. The overall processing process is shown in Figure 3.

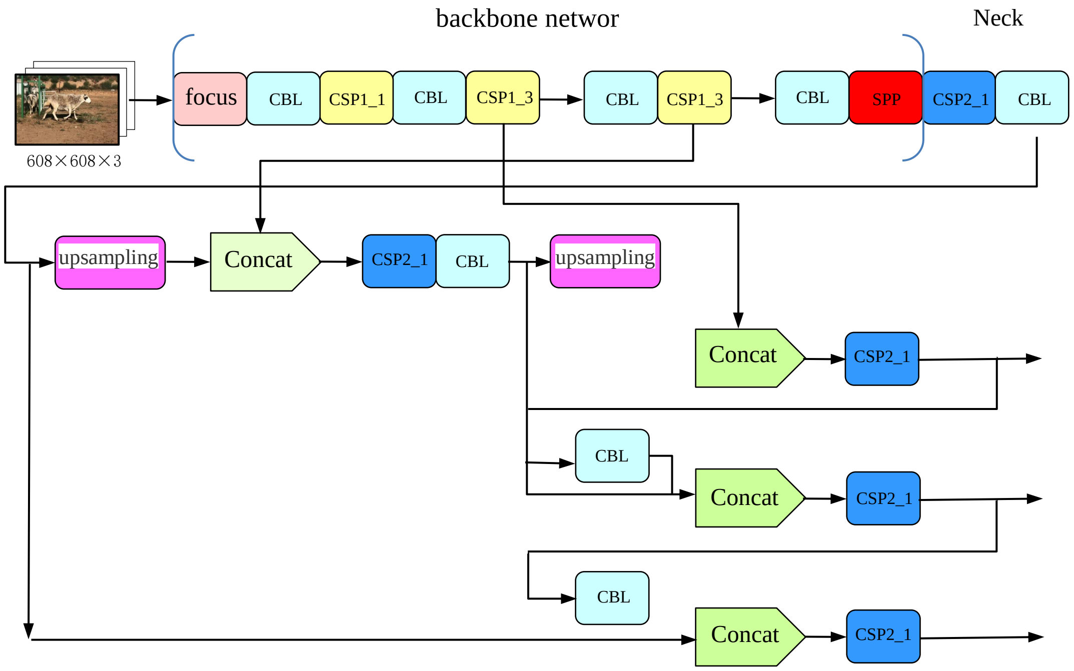


Fig. 3 - Input sheep graph in YOLOv5 backbone network and neck network processing process

In the detection process, the sheep diagram needs to be sliced first. The image of 608×608×3 is changed into 304×304×12 after a slicing operation, and then a convolution process with a convolution kernel of 32 is carried out. Finally, the feature image of 304×304×32 will be changed.

The slicing process is shown in Figure 4.

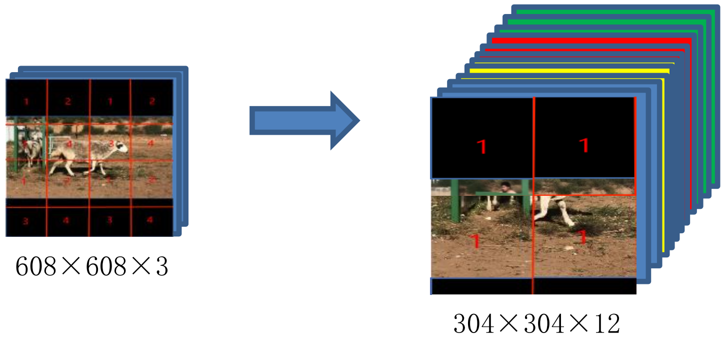


Fig. 4 - Image slicing process

CBL consists of convolutional layer (CNN), normalized processing layer (BN) and activation layer (Relu), which is used to extract sheep image features. SPP consists of convolution layer and pooling layer, which further extracts features and simplifies computation to improve detection speed.

YOLOv4 and YOLOv5 add CSP algorithm to the backbone network to increase the learning ability of CNN (convolutional neural network). It also reduces computing bottlenecks and memory costs. Through the trunk network and then into the neck network, the neck network is a clustering combination of features extracted from the trunk network. The neck Network of YOLOv4 adopts the Path Aggregation Network (PANet) structure.

In Figure 5, PANet uses feature pyramid and path aggregation technology to transmit information from the bottom layer to the top layer. FPN combines the features of the upper layer from top to bottom by up-sampling. Finally, the predicted feature graph is obtained.

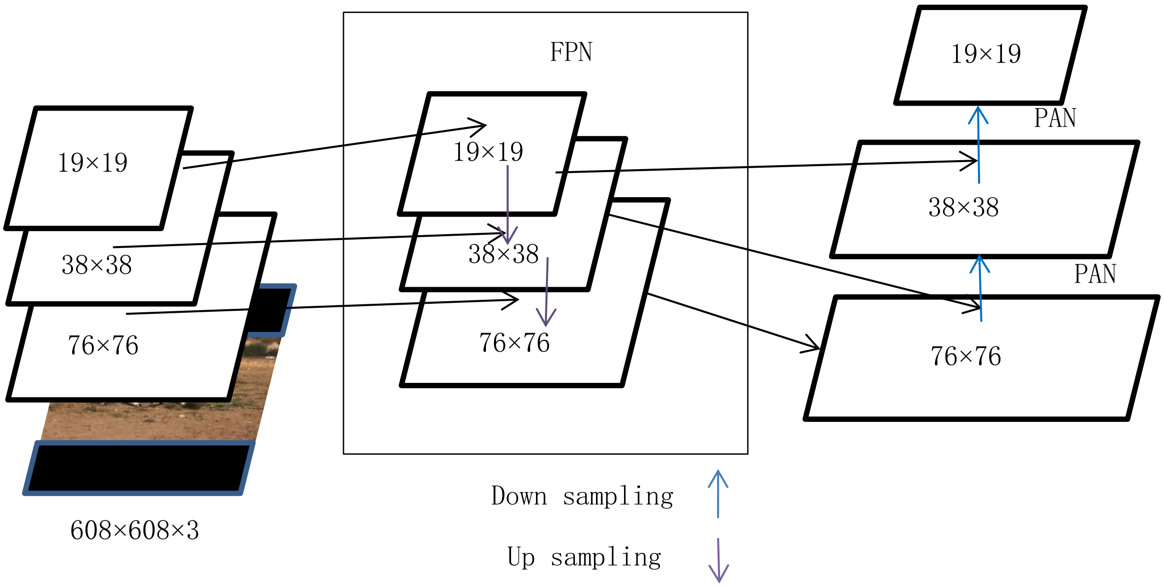


Fig. 5 - Extraction feature and aggregation feature

In the figure, the red box represents Ground truth box and the yellow box represents Prediction box.

IOU = Red box ∩ yellow box/Red box ∪ yellow box.

The blue part of the figure is the difference set. On the output side, boundingBox's loss function GIoU needs to be computed.

$$GIoU = IoU - \frac{|\text{Difference set}|}{c} \quad (1)$$

$$GIoU\_loss = 1 - GIoU \quad (2)$$

The process is shown in Figure 6.

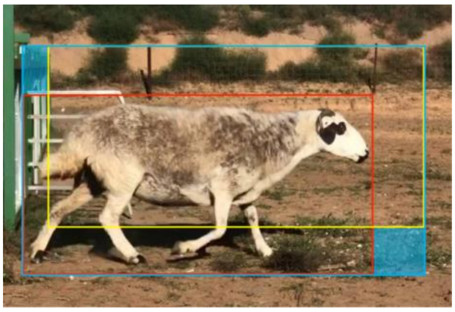


Fig. 6 - Calculation method of Glou_loss

Finally, multiple prediction boxes are screened, and NMS non-maximum box suppression is used to clear the boxes with low confidence, leaving the prediction boxes with the highest confidence.

Dash line counting program logic

Because sheep can be divided into two cases of exit and return when crossing the channel, the bidirectional collision counting method is adopted in this paper. There are two lines to control the total. The process is shown in Figure 7.

Firstly, the corresponding model was imported to detect the target video, and the sheep in the view Angle were randomly assigned numbers and tracked. In this case, you need to set three lists: blue line list, yellow line list and numbered list. The sheep are numbered and encapsulated into a class. The number and the number of line collisions are added into the class to judge whether the numbered sheep has been counted. When the number of collisions is greater than or equal to 2, it will be automatically cleared from the three lists, so as to avoid repeated counting and achieve the purpose of accurate counting. Setting the position of yellow and blue lines at the same time can realize the effect of counting sheep in different directions.

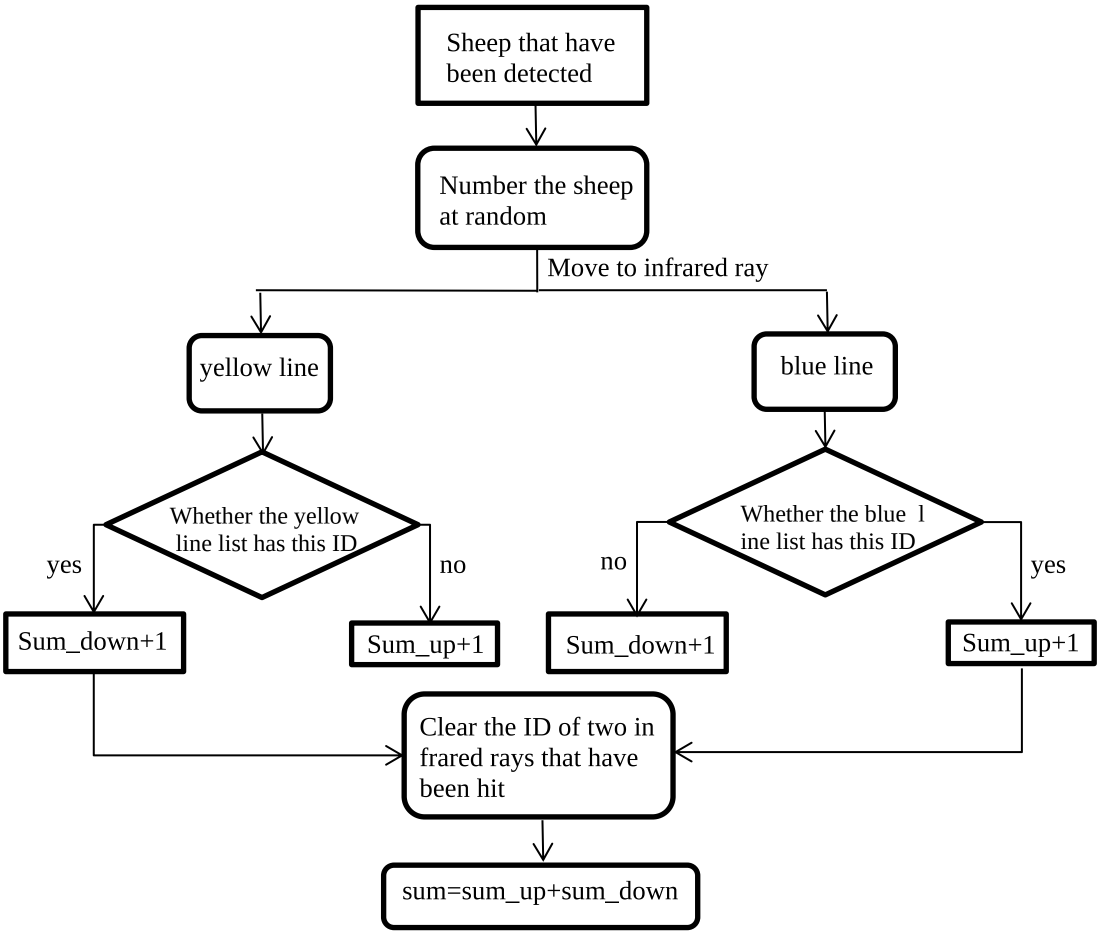


Fig. 7 - Program flow of line collision counting

DESIGN OF SHEEP COUNTING DETECTION SYSTEM

System overview

Sheep is the most important product of animal husbandry, and the automatic management of sheep can improve the efficiency of breeding. Nowadays, many herdsmen raise hundreds of sheep, so the problem of sheep counting becomes an important task in sheep management. The sheep management system described in this paper is based on YOLOv5 algorithm, and each sheep is numbered by using DeepSORT tracking monitoring method. Finally, the statistical function of the target sheep is completed by the method of line bumping. The flow chart of sheep counting is shown in Figure 8.

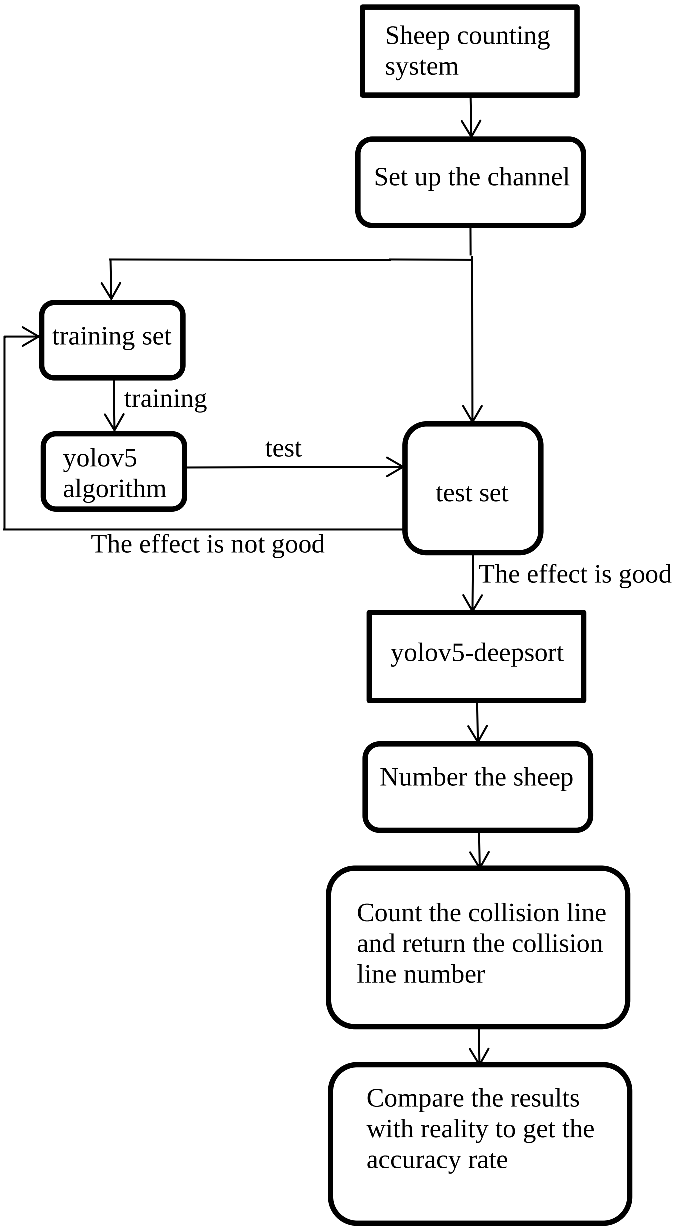


Fig. 8 - Sheep counting flow chart

Monitoring platform construction

The monitoring platform uses a PC, a xiaomi router 4A, a Hikvision B12V2-I/ POE 4mm and other experimental equipment in the barn of Helinger County Beiqi Pharmaceutical Equipment Company. Based on Visual Studio Code -- Python programming environment, using NVIDIA GeForce GTX 960M graphics card (4G memory). Object detection was carried out in the environment of Windows10 operating system, Cuda version Cuda10.0 and network framework Pytorch, and deep learning was carried out using YOLOv5 algorithm as the framework. The platform device is shown in Figure 9.

Fig. 9 - Experimental field device

SYSTEM TESTING AND ANALYSIS

Sheep data training process

During data training, the data collected in Section 1.1 were marked, and sheep with four postures were marked as sheep_body. After image screening and cleaning, 640 images were selected to construct data for training. These images were constructed according to Pascal VOC data set format, marked with Labelimg and generated corresponding XML files, which were generated into corresponding coordinate TXT files. The next step is to divide these data images into training sets and test sets. Trainval_percent = 0.9 is selected in this paper, that is, the training set is 0.9, and the rest is the division method of test set. The training parameters were epochs=50 and the initial learning rate α =0.001. The corresponding best. Pt file was obtained after the training, which was the sheep detection model.

Monitoring system implementation

The realization of sheep counting system needs to build experimental environment first. Due to the habit of timid gathering, sheep like to gather together, so the sheep in the video are stuck together, which makes target detection and number statistics impossible. Therefore, a channel is needed for sheep to queue through, so as to solve the problem of sheep sticking and unable to separate. Therefore, a channel is constructed. This channel is a right Angle channel, connecting the entrance and exit of the sheepfold. It can ensure that sheep will walk along the channel when they leave and return to the circle, greatly reducing the difficulty of target detection. Figure 10 is the channel diagram used in the experiment.

Fig. 10 - Experimental channel

System experiment process and analysis

The overall situation of the system experimental process and training process is shown in Table 1.

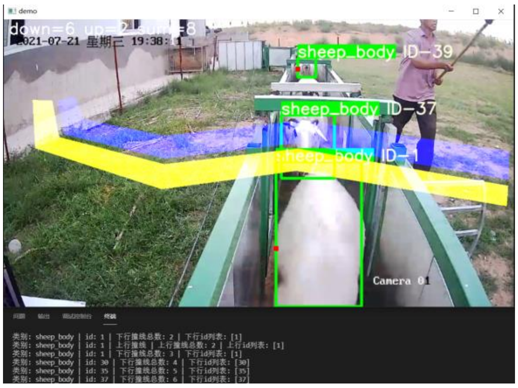
Table 1

Experiment and training process of the system

The experimental process	The training process
First make a passage that allows sheep to pass through one by one. The camera is installed on the beam at the end of the channel, and then the training data is collected. The data set needs to be clearly visible to ensure that the sheep under test do not stick to each other and have a certain distance from each other.	The collected data of training set is extracted frame by frame, and then extracted frame is saved. Then, data cleaning was carried out, and the pictures with distortion, no sheep and unclear sheep were deleted. 640 pictures were screened out for training. Then write two script files to make data sets, and finally train on GPU.

After analyzing the experimental results, the system counted the number of sheep in different directions of the passage, and the passage was arranged along the north-south direction of the pasture. Therefore, when the system runs, sheep counting lines are configured in the north-south and east-west directions respectively. The sheep are counted as they pass the count line. The real-time running status of the experiment is shown in Figure 11. During the experiment, the number of sheep passing through the channel can be detected in real time in both north-south and east-west directions, and the system is verified to run stably.

East West counting



North South counting

Fig. 11 - Real-time experimental results graph

CONCLUSIONS

In this paper, a target detection method based on machine learning is proposed for flock counting. YOLOv5 is adopted as the core algorithm, and the counting function is optimized and analyzed in the real scene of pasture. Experiments show that this method can accurately record the number of sheep, and the system runs stably and reliably. It can be used as an extension module of sheep related management system, or as a separate device for sheep management. This scheme provides a practical and feasible scheme for the design of sheep counting device.

ACKNOWLEDGEMENT

This research, titled 'A REAL-TIME SHEEP COUNTING DETECTION SYSTEM BASED ON MACHINE LEARNING', was funded by the research project of Scientific Research Start of Doctor of Shanxi Agricultural University (Research on Detection Technology of Agricultural Internet of Things Model, 2017YJ30). The authors are grateful and honored to receive this support.

REFERENCES

[1] Anlan D., Xuhui Z., Xiuguo Z., et al. (2019), A novel method for the group characteristics analysis of yellow feather broilers under the heat stress based on object detection and transfer learning.[J]. *R and D National Institute for Agricultural and Food Industry Machinery - INMA Bucharest 3*.

[2] Bochkovskiy A., Wang C. Y., Liao H. Y. M., (2020), Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934.

[3] Chen Y., Li W., Sakaridis C., et al. (2018), Domain Adaptive Faster R-CNN for Object Detection in the Wild// 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *IEEE*.

[4] Chen M., Yang H. Z., (2019), A Method with Adaptive Seed Point Substitution for Counting Overlapped Cell. *Control Engineering of China*, 26, 02, 236-240.

[5] Deng, X., Yan, X., Hou, Y., et al. (2021). DETECTION OF BEHAVIOUR AND POSTURE OF SHEEP BASED ON YOLOv3. *INMATEH-Agricultural Engineering,* 64(2).

[6] Girshick R., (2015), Scale-aware Fast R-CNN for Pedestrian Detection. Computer Science.

[7] Guo L., Bian G. Q., (2019), Study on library access flow forecast based on bat algorithm optimizing relevant vector machine. *Modern Scientific Instruments*, 1, 184-188.

[8] Jiang J. G., Wang T., Qi M. B., et al. (2012), Statistical algorithm for traffic flow based on ViBe. *Journal of Electronic Measurement and Instrument*, 26, 006, 558-563.

[9] Jia W., Xu S., Liang Z., et al. (2021), Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. *IET Image Processing.nic world*, 03, 77-78+81.

[10] Liu C., Li X., Li Q., et al. (2021), Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model. *Neurocomputing*, 430, 174-184.

[11] Lu J. G., Zhang L., (2019), Crowd counting model based on multi-scale multi-column convolutional neural network. *Journal of Computer Applications*, 39, 12, 3445-3449.

[12] Redmon J., Divvala S., Girshick R., et al. (2016), You Only Look Once: Unified, Real-Time Object Detection. *IEEE*.

[13] Redmon J., Farhadi A., (2017), YOLO9000: Better, Faster, Stronger// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 6517-6525.

[14] Redmon J., Farhadi A., (2018), Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767.

[15] Song Y., Cui P. Y., Ju H. H., (2006), A Computational Optimization Method of the SSD and NCC Algorithms of the Stereo Matching. Computer Engineering and Applications, 42, 2, 42-44.

[16] Sun X., Wu P., Hoi S., (2018) Face Detection using Deep Learning: An Improved Faster RCNN Approach. *Neurocomputing*, 299, JUL.19, 42-50.