Deep Learning-Based Autonomous Cow Detection for Smart Livestock Farming

Yongliang Qiao[1], Yangyang Guo[2(✉)], and Dongjian He[3]

[1] Australian Centre for Field Robotics (ACFR), Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia

[2] School of Internet, Anhui University, Hefei 230039, Anhui, China
guoyangyang113529@ahu.edu.cn

[3] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, Shaanxi, China

Abstract. Animal sourced protein is increasing rapidly due to the growing of population and incomes. The big data, robots and smart sensing technologies have brought the autonomous robotic system to the smart farming that enhance productivity and efficiency. Therefore, a YOLOv4-SAM was proposed to achieve high detection precision of cow body parts in long-term complex scenes. The proposed YOLOv4-SAM consists of two components: YOLOv4 is for multi-scale feature extraction, while the Spatial Attention Mechanisms (SAM) highlights the key cow biometric-related features. By doing this, visual biometric feature representation ability is enhanced for improving cow detection performance. To verify the performance of YOLOv4-SAM, a challenging dataset consisting of adult cows and calves with complex environments (e.g., day and night, occlusion, multiple target) was constructed for experimental testing. The precision, recall, mIoU and of the proposed YOLOv4-CBAM were 92.29%, 96.51%, 77.22% and 93.13%, respectively. The data shows that its overall performance was better than that of the comparison algorithm (Faster R-CNN, RetinaNet and YOLOv4). In addition, object detection based on the YOLOv4-SAM model can capture the key biometric-related features for cow visual representation and improve the performance of cow detection. In addition, the detected height difference between head and leg proved the capability in automatic identification of lame cows. The proposed deep learning-based cow detection approach provides a basis for developing an automated system for animal monitoring and management on commercial dairy farms.

Keywords: Autonomous detection · YOLOv4 · Attention mechanism · Deep learning · Precision livestock farming

1 Introduction

Precision farming is key for producing more food for increasing world population, especially the livestock production that provides valuable protein [1]. To meet human demand for animal products, the animal production needs to improve its efficiency and operations for higher productivity [2]. In the past decades, intelligent mechanical equipment

C. Yu et al. (Eds.): GPC 2022, LNCS 13744, pp. 246–258, 2023.
https://doi.org/10.1007/978-3-031-26118-3_19

and robotics have been one of the main frame-works which focusing on minimizing environmental impacts and simultaneously maximizing agricultural productivity [3, 4]. Automated systems with smart sensors, big data processing based on artificial intelligence, and robots can be used to enhance the efficiency of production such as crop harvesting, fruit picking and animal farming [5, 6].

In smart agriculture, the automatic, efficient and accurate acquisition of animal information has been a key prerequisite. The vision-based animal monitoring, body parts detection, and welfare evaluation can be completed automatically without stress [2]. More recently, deep learning can more effectively realize visual feature-based object detection and behavior recognition with its network depth and feature representation ability [7]. Riekert et al. [11] combined Faster R-CNN and Neural Architecture Search (NAS) to construct a pig position detection and pose recognition system, and verified the feasibility of their methods. Shao et al. [12] designed a cattle detection system based on Convolutional Neural Networks (CNNs), obtained a precision of 0.957. Jiang et al. [13] proposed FLYOLOv3 method to detect body parts (e.g., head, leg and trunk) of individual dairy cow. Although the above approaches demonstrated the feasibility of deep learning-based animal detection, it is still challenging to achieve accurate detecting of key areas of cow in a complex farm environment (e.g., multi-cows, occlusion, different illumination, day and night) [14].

More recently, YOLO network has been a popular method for object detection [18]. In YOLO family, YOLOv4 is better than YOLOv1, YOLOv2 and YOLOv3, and has excellent speed and accuracy [19]. YOLOv4 is an end-to-end real-time deep learning-based method of object detection proposed in 2020, and it has been proved with high detection accuracy and speed on many datasets [19]. In addition, the attention mechanism can improve the extraction ability of target-related features and reduce the interference of non-target area features, thereby improving the effect of target detection models and is widely used [20, 21].

To improve remote animal detection accuracy, we proposed YOLOv4-SAM model to detect the cow. Firstly, YOLOv4 was used to extract multi-scale features from sample images. Then, Spatial Attention Mechanisms (SAM) was used to highlight the animal biometric-related features and enhance the animal detection performance. In this study, images of dairy adult cows and calves during the day and night were acquired for testing. In our detection experiments, whole animal, head, and four legs were detected respectively. And a further qualitative analysis of lame recognition was conducted based on height difference of detected legs and head.

The contributions of this paper include: (1) We integrated the SAM attention mechanism into YOLOv4 and proposed YOLOv4-SAM based approach for cow boy part detection. The proposed YOLOv4-SAM improves the biometric feature representation ability and enhances cow body part detection performance; (2) A long-term challenging dataset consisting of adult cow and calf with complex environments (e.g., day and night, occlusion, multiple target) was constructed for detection verification. The results indicate that the YOLOv4-SAM method was superior to the comparison method, with high detection accuracy and fast processing speed, which satisfy the real-time requirement of smart farm applications; and (3) Th height difference of YOLOv4-SAM based leg and head detection could further used to detect lameness. Overall, the proposed

YOLOv4-SAM provides a real-time and high accuracy cow body part detection approach in complex scenes, which facilitates long-term autonomous animal detection and health/welfare evaluation.

2 Material and Methods

2.1 Data Acquisition

In this experiment, the data sets for calf and adult cattle were sourced from Yangling Keyuan Cloning Co., Ltd. China. For calf data, a Holstein calf was monitored in a rectangular enclosure (4 m × 2 m × 1.5 m). The height of the installed camera was 0.75 m and the distance from the fence is 2.5 m. For adult cow data, the camera was placed on a support beam of the 1.8-m-high feed shed, and it was 35 m away from the aisle. Camera setup for calf and cow video recording is presented in Fig. 1.

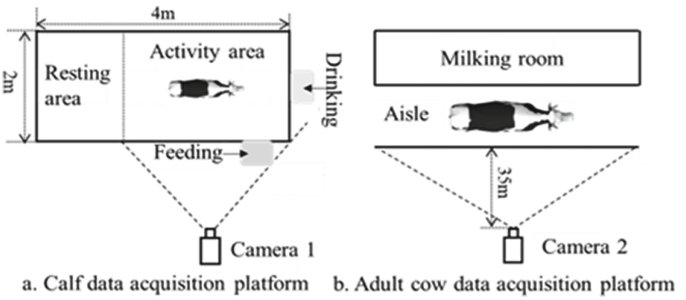

Fig. 1. Video acquisition setup diagram of cattle.

The dataset acquired consider is really challenging due to the factors: multi-cow appeared and exist occlusion; Lighting is changing from dawn to dusk; the complex background including crush, soil ground, building background and so on. The frame rate, bit rate and resolution of the captured cattle videos were 25 fps, 2000 kbps, 704 pixels × 576 pixels respectively. A total of 1040 images were obtained, including 540 images of adult cows and 500 images of calf. Among these images, 600 and 440 images were randomly selected as the training and testing datasets. In our experiments, the whole cow, the head, and legs (left front leg, left hind leg, right front leg, and right hind leg) were labeled manually with bounding boxes, respectively, for key body areas detection testing.

2.2 YOLOV4-SAM Model for Detecting Dairy Cow

In order to improve the detection performance for key cow parts, we integrated SAM attention module into the YOLOv4 object detector, which had higher accuracy in cow detection and monitoring.

The proposed YOLOv4-SAM approach extracted multi-scale features using YOLO4 object detector and learning feature importance through SAM, enhancing the representation of animal biometric related features and improving the animal detection performance. The proposed YOLOv4-SAM-based detection method includes four parts, backbone, neck, SAM, and head as shown in Fig. 2.

- CSPDarknet53 as backbone network of YOLOv4 algorithm was be used to extract target features [19].
- The Neck was composed of the PANet and SPPNet. PANet proposed a bottom-up information propagation path enhancement method, which realized bottom-up feature extraction through convolution and up-sampling, and realized top-down feature extraction through down-sampling, thereby better fusing the extracted features [22]. The SPPNet module can concatenate feature maps from different core sizes together as an output, effectively increasing the backbone's acceptance domain and separating significant context features [23].
- SAM was used to enhance the weight of the target candidate region after the convolution block [24].
- And then the YOLOv3 head was used to realize object detection [25].

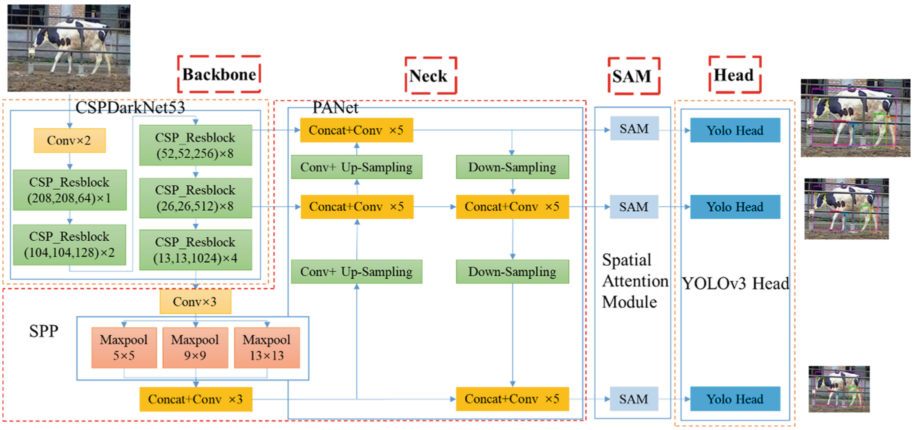

Fig. 2. YOLOv4-SAM based cow detection framework.

2.2.1 SAM Module for Feature Optimization

The SAM mainly encoded spatial pixel correlations in the feature map into local features to enhance their representation ability. The SAM module structure is illustrated in Fig. 3. In the feature map $F_t \in R^{C \times H \times W}$, C, H and W represent the channel number, height, and width, respectively. Pooling operator includes the max pooling and the average pooling. The equation is as follows:

$$P_a = AvePool(F_t) \tag{1}$$

$$P_m = MaxPool(F_t) \tag{2}$$

where $AvePool(\cdot)$ and $MaxPool(\cdot)$ indicate average pool and max pool, respectively.

Then P_a and P_m form a new feature P_{am} by *Concat()* function. After that, the spatial attention map *As* is obtained by the *Conv ()* function and the *Sigmoid* $(\cdot)$ function:

$$A_s = Sigmoid(Conv(Concat(P_a, P_m))) \tag{3}$$

Finally, the feature map Fs $\in$ $R^{C\times H\times W}$ can be calculated as:

$$F_s = A_s \otimes F_t \tag{4}$$

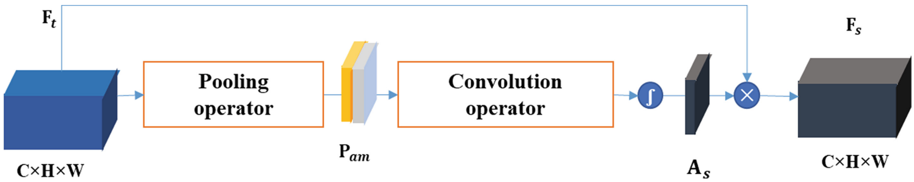

Fig. 3. The SAM module structure.

2.3 Loss Function

The loss function based on YOLOv4-SAM model consists of bounding box location loss function (L_{CIoU}), confidence loss function (L_{conf}) and classification loss function (L_{class}), which was used in the cow detection model based on YOLOv4-SAM. If there is no target, only the L_{conf} is calculated. If there is a target, L_{CIoU}, L_{conf} and L_{class} are calculated. The loss function equations are as follows [26]:

$$L_{CIoU} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}\left[1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v\right] \tag{5}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{6}$$

$$\alpha = \frac{v}{((1 - IoU) - v)} \tag{7}$$

$$L_{conf} = -\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}\left[\hat{C}_i \log(C_i) + \left(1 - \hat{C}_i\right)\log(1 - C_i)\right] - \lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{noobj}\left[\hat{C}_i \log(C_i) + \left(1 - \hat{C}_i\right)\log(1 - C_i)\right] \tag{8}$$

$$L_{class} = -\sum_{i=0}^{S^2} I_{i,j}^{obj} \sum_{c \in classes} \left[\hat{P}_i(c) \log\left(P_i(c)\right) + \left(1 - \hat{P}_i(c)\right) \log\left(1 - P_i(c)\right) \right] \tag{9}$$

$$Loss = L_{CIoU} + L_{conf} + L_{class} \tag{10}$$

where, S is the number of grids, B is the number of prior boxes in each grid. λ_{noobj} represents weight. $I_{i,j}^{obj}$ and $I_{i,j}^{noobj}$ are used to determine whether the $j-th$ priori box of the $i-th$ grid contains the object. If yes, $I_{i,j}^{obj}$ is 1 and $I_{i,j}^{noobj}$ is 0. Otherwise $I_{i,j}^{obj}$ is 0 and $I_{i,j}^{noobj}$ is 1. IoU refers to the ratio of intersection and union of the prediction and actual bounding boxes. ρ *()* is the Euclidean distance, and c is the diagonal distance between the predicted box and the closure area of actual box. b, w and h are the center coordinates, width and height of the prediction box, respectively. b^{gt}, w^{gt} and h^{gt} are the center coordinates, width and height of the actual box, respectively. α is weight factor, v is similarity ratio of length to width. C_i means the confidence of prediction and label box. $P_i(c)$ is the classification probability of different categories; c is the number of detection categories.

3 Experimental Setup

3.1 The Used Dataset

The sample descriptions of training and testing sets are shown in Table 1.

Table 1. Datasets description in the experiments

Dataset	Number of datasets
Train	Adult cow: 300 (270 days; 30 nights) Calf: 300 (270 days; 30 nights)
Test	Adult cow: 240 (220 days; 20 nights) Calf: 200 (180 days; 20 nights)

3.2 Network Training Platform

In our work, all data analysis works were carried out on a computer equipped with a GeForce GTX 1080 Ti GPU and I9-7920X CPU@2.9 GHz, using Keras framework. To verify the effectiveness of the dairy cow key parts detection model, Faster R-CNN [27], RetinaNet [28] and YOLOv4 [19] were used for comparison. In addition, the input size of all networks was 416 × 416 × 3, epoch was 1000, batch size was 16, and learning rate was 0.0013. All the initial weights of the network were random. The other parameters were the default settings.

4 Results

4.1 Detection Performance Comparison

The proposed YOLOv4-SAM based cow detection was compared to other CNN based approaches in Table 2. It can be seen that the precision and recall of the YOLOv4-SAM approach is up to 92.29% and 96.51%, respectively, which is higher than that of Faster R-CNN (53.85%,63.25%), RetinaNet (74.95%,75.69%) and YOLOv4 (91.86%,96.51%). For the mIoU, YOLOv4-SAM method is 77.22%, higher than that of Faster R-CNN (51.36%), RetinaNet (76.36%) and YOLOv4 (75.18%). In addition, it can be noticed that the mAP@0.5 of YOLOv4-SAM is 93.13%, which is higher than that of YOLOv4 (93.08%). These results demonstrated that the proposed YOLOv4-SAM approach can pay more attention to the visual features related to the animal body and enhance the animal detection ability.

In addition, the detection speed of our proposed YOLOv4-SAM (40 f/s) is fast than that of Faster-RCNN and RetinaNet networks, and lower than the original YOLOv4 network. But the average recognition speed could reach 40 FPS, which would highly possible to satisfy the real-time requirement (>20 FPS) of robotic based autonomous cow detection. Overall, the proposed YOLOv4-SAM highlights the related animal detection related features and enhances the detection performance, which provides a favorable solution for the remote animal detection in smart livestock farming.

Table 2. Comparison of different GDM methods.

Method	Precision (%)	Recall (%)	mIoU (%)	mAP@0.5 (%)	FPS (f/s)
Faster R-CNN	53.85	63.25	51.36	59.72	27
RetinaNet	74.95	75.69	76.36	74.38	29
YOLOv4	91.86	96.51	75.18	93.08	55
YOLOv4-SAM	**92.29**	**96.51**	**77.22**	**93.13**	**40**

Cow and calf images from the different scenes (e.g., day and night, occlusion, multiple target) were selected to evaluate the performance of the model. Figure 4 shows the detection results of our proposed YOLOv4-SAM approach. It shows that the YOLOv4-SAM approach could recognize the images of cow and calf at night, and accurately detect the head and legs, which is beneficial to the long-term cattle detection in smart animal husbandry. All this indicating that robustness of the proposed YOLOv4-SAM approach.

Fig. 4. Examples of YOLOv4-SAM based cow and calf detection. In the detection result, the orange, green, purple, light green, yellow, red and blue boxes are adult cow area, claf area, head, outer front leg, inner front leg, outer hind leg and inner hind leg, respectively. (Color figure online)

4.2 Detection Results of Key Body Parts of Cattle

To further analysis the detection performance of different body parts, in Table 3, the numbers of detected body parts (including tp and fp), precision, recall, IoU and mAP@0.5 are presented.

The correctly detected number (tp) and false detected number (fp) of YOLOv4-SAM are 2759 and 216 respectively, and the tp is 2757 and fp is 260 in YOLOv4. The overall precision, recall, and map@0.5 of YOLOv4-SAM are 92.29%, 96.51% and 93.13% respectively, which higher than that of YOLOv4. In animal detection, the proposed YOLOv4-SAM achieved 96.95% and 99.00% precision for the whole body of adult cow and calf, respectively. As the animal body truck is large than other body parts (e.g., head, leg), both YOLOv4 and YOLOv4-SAM achieved higher detection accuracies for animal itself (e.g., adult cow and calf), and the detection recall of the YOLOv4-SAM approach is more noticeable. From these data, it is clear that the SAM module in YOLOv4 improves the animal visual biometric feature (e.g., shape, contour, and coat color) representation ability, enhancing the cow detection in different environments.

Table 3. Comparison of different body part detection performance.

Method	Species	GT	Detect (tp)	Detect (fp)	Precision (%)	Recall (%)	IoU (%)	mAP@0.5 (%)
YOLOv4	Adult cow	286	286	9	96.95	100.00	85.75	99.94
	Calf	199	199	6	97.07	100.00	89.02	99.68
	Head	481	471	11	97.73	98.54	80.19	90.91
	Outer front leg	482	460	47	90.73	95.44	72.04	90.63

(*continued*)

Table 3. (*continued*)

Method	Species	GT	Detect (tp)	Detect (fp)	Precision (%)	Recall (%)	IoU (%)	mAP@0.5 (%)
	Inter front leg	472	454	36	92.65	96.19	73.38	90.57
	Outer hind leg	479	456	67	87.19	95.20	71.06	90.51
	Inner hind leg	465	431	84	83.69	92.69	66.44	89.36
	All	**2855**	**2757**	**260**	**91.86**	**96.51**	**75.18**	**93.08**
YOLOv4-SAM	Adult cow	286	286	9	96.95	100.00	90.52	99.87
	Calf	199	199	12	99.00	100.00	92.62	99.82
	Head	481	476	8	98.35	98.96	82.15	90.91
	Outer front leg	482	458	33	93.28	95.02	75.00	90.51
	Inter front leg	472	460	34	93.12	97.46	75.16	90.67
	Outer hind leg	479	454	61	88.16	94.78	73.85	90.52
	Inner hind leg	465	435	69	86.31	93.55	67.75	89.58
	All	**2864**	**2759**	**216**	**92.29**	**96.51**	**77.22**	**93.13**

In addition, the detection results of Table 3 show that the precision of the YOLOv4-SAM for head (98.35%), outer front leg (93.28%), inter front leg (93.12%), outer hind leg (88.16%) and inner hind leg (86.31%) is higher than that of YOLOv4 (97.73% for head, 90.73% for outer front leg, 92.65% for inter front leg, 87.19% for outer hind leg and 863.69% for inner hind leg). It also can be noticed that head detection precision is higher than that of leg, the main reason is that cow head account for a large area and not occluded by other body parts, thus the extracted visual features from head has few disturbing factors. All these increased values show that the application of SAM was feasible in the cow body parts' detection.

4.3 Application of Lame Detection in Dairy Cows

Cow lameness can reflect the health problems of cattle, and the detection of abnormal behavior is important for farms [29]. When a cow is lame, its head position will be lower than that of a normal cow.

To further explored the relative height of the head and leg of the lame cow and the normal cow. Firstly, 30 images of normal cows and lame cows were selected, respectively. Then, yolov4-SAM model was used to detect the head and legs of cows. The *y* in the central coordinate (x, y) of the bounding box of the head was used to represent the height information of the head. Similarly, y_i in the central coordinate (x_i, y_i) of the bounding box of the leg was obtained, i represented the number of legs and the average value y_{avr} of y_{i} was taken as the height information of the legs. The difference between y and y_{avr} was denoted as relative height of the head position and the leg position (Fig. 5).

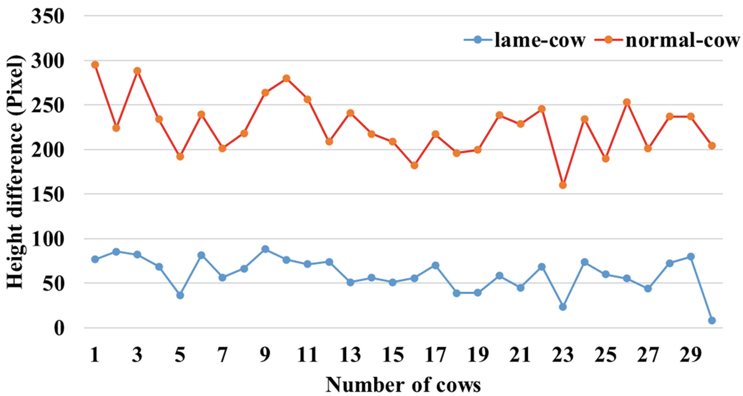

Fig. 5. The relative height difference between the head and leg of normal cows and lame cows.

It can be seen from Fig. 5 that the height difference between the head and leg of lame cows (average is 60.6 pixel) is lower than that of normal cows (average is 165.9 pixel). That means using the position of cow head and leg, namely, the height difference, cow lameness behavior could be detected.

5 Discussion

5.1 Impact of Object Size and Number on Detection Performance

As the YOLOv4 using bounding box to detect object, detecting object with different sizes could have varying performance. The whole-cow body account for a large proportion result in high detection accuracy, while the leg area or head accounted for the small proportion obtained lower detection precision. As illustrated in Table 3, the whole cow detection accuracy was higher than that of head and leg. Although there was a detection difference between whole-cow body, head, and leg, the overall detection performance was over 90% due to the multi-scale feature extraction ability in YOLOv4-SAM.

In Fig. 4, when there is a part of the cow in the scene; or when multiple cows occlude each other, the proposed YOLOv4-SAM could still well detect. This is because that multi-scale feature extraction ability of YOLOv4 and the key feature highlighting mechanism of SAM attention module, visual biometric related features could be well extracted for cow detection.

5.2 Impact of Different Scenes on Detection Performance

To verify the robustness of the YOLOv4-SAM method, the experimental samples contain data from different scenes, as shown in Fig. 4. It shows that the whole body, head and legs of adult cows and calves in different scenes (e.g., day and night, occlusion, multiple target) are well detected. In addition, the adult cow samples include railings, which will occlude part of the legs. In this case, the YOLOv4-SAM method can still detect the legs well, but due to the loss of some information, it will also cause misrecognition. From Table 2, in the sample data set containing multiple scenarios, the performance of the YOLOv4-SAM in this study is better than other comparison methods (Faster R-CNN, RetinaNet and YOLOv4).

5.3 Analysis of YOLOv4-SAM Model Applicability

YOLOv4 model is a real-time object detection algorithm, which integrated the characteristics of YOLOv1, YOLOv2, YOLOv3 and other advanced methods, and improved the detection speed and detection accuracy of the object detection network. And the attention mechanism SAM can highlight the animal biometric-related features to enhance the animal detection performance. So, the YOLOv4-SAM model not only retains the performance of YOLOv4, but also improves the recognition ability of the network. In addition, this study further explored the identification of lame cows. It can be seen from Fig. 5 that the head-to-leg height distance of lame cows is lower than that of normal cows, which can be used as a basis for judging the lame cow. This model can also be applied to assist UAV or UGV for monitoring group animals or animal body parts (e.g., heads, legs, and back, etc.). The movement information of an individual animal's head and legs can be further detected for analyzing animal behavior.

6 Conclusions

In this study, a YOLOv4-SAM based approach was proposed to detect cow and its different body parts remotely. The proposed approach integrating attention mechanism SAM into YOLOv4, enhancing animal visual biometric feature representation ability, which provides a new way for long-term and real-time animal detection for further autonomous based smart livestock farming. A challenging cow dataset consisting of adult and calf with complex environments (e.g., day and night, occlusion, multiple target) was constructed for experimental testing. The proposed YOLOv4-SAM based approach outperformed Faster R-CNN, RetinaNet, and YOLOv4. Meanwhile, the detection performance of different cow body was investigated. Experimental results demonstrated that the proposed YOLOv4-SAM approach could capture key biometric-related features for animal visual representation, improving the performance of cow detection. In addition, the detected height difference of head and leg could be further used to identify lame cow from the health group. Overall, the proposed deep learning-based cow detection approach is favorable for autonomous cow management in smart livestock farming.

References

1. Qiao, Y., et al.: Intelligent perception for cattle monitoring: a review for cattle identification, body condition score evaluation, and weight estimation. Comput. Electron. Agric. **185**, 106143 (2021)
2. Fournel, S., Rousseau, A.N., Laberge, B.: Rethinking environment control strategy of confined animal housing systems through precision livestock farming. Biosys. Eng. **155**, 96–123 (2017)
3. Mishra, S., Syed, D.F., Ploughe, M., Zhang, W.: Autonomous vision-guided object collection from water surfaces with a customized multirotor. IEEE/ASME Transactions on Mechatronics **26**, 1914–1922 (2021)
4. Lim, J., Pyo, S., Kim, N., Lee, J., Lee, J.: Obstacle magnification for 2-D collision and occlusion avoidance of autonomous multirotor aerial vehicles. IEEE/ASME Trans. Mechatron. **25**(5), 2428–2436 (2020)
5. Zhang, Z., Kayacan, E., Thompson, B., Chowdhary, G.: High precision control and deep learning-based corn stand counting algorithms for agricultural robot. Auton. Robot. **44**(7), 1289–1302 (2020). https://doi.org/10.1007/s10514-020-09915-y
6. Ding, H., Gao, R.X., Isaksson, A.J., Landers, R.G., Parisini, T., Yuan, Y.: State of AI-based monitoring in smart manufacturing and introduction to focused section. IEEE/ASME Trans. Mechatron. **25**(5), 2143–2154 (2020)
7. Qiao, Y., Truman, M., Sukkarieh, S.: Cattle segmentation and contour extraction based on mask R-CNN for precision livestock farming. Comput. Electron. Agric. **165**, 104958 (2019)
8. Beggs, D., Jongman, E., Hemsworth, P., Fisher, A.: Lame cows on australian dairy farms: a comparison of farmer-identified lameness and formal lameness scoring, and the position of lame cows within the milking order. J. Dairy Sci. **102**(2), 1522–1529 (2019)
9. Okinda, C., et al.: A review on computer vision systems in monitoring of poultry: a welfare perspective. Artif. Intell. Agric. **4**, 184–208 (2020). https://doi.org/10.1016/j.aiia.2020.09.002
10. Li, G., et al.: Practices and applications of convolutional neural network- based computer vision systems in animal farming: a review. Sensors **21**(4), 1492 (2021)
11. Riekert, M., Klein, A., Adrion, F., Hoffmann, C., Gallmann, E.: Automatically detecting pig position and posture by 2D camera imaging and deep learning. Comput. Electron. Agric. **174**, 105391 (2020)
12. Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., Naemura, T.: Cattle detection and counting in UAV images based on convolutional neural networks. Int. J. Remote Sens. **41**(1), 31–52 (2020)
13. Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., He, D.: FLYOLOv3 deep learning for key parts of dairy cow body detection. Comput. Electron. Agric. **166**, 104982 (2019)
14. He, D., Liu, D., Zhao, K.: Review of perceiving animal information and behavior in precision livestock farming. Trans. Chin. Soc. Agric. Mach **47**(5), 231–244 (2016)
15. Mahmud, M.S., Zahid, A., Das, A.K., Muzammil, M., Khan, M.U.: A systematic literature review on deep learning applications for precision cattle farming. Comput. Electron. Agric. **187**, 106313 (2021)
16. Meena, S.D., Agilandeeswari, L.: Smart animal detection and counting framework for monitoring livestock in an autonomous unmanned ground vehicle using restricted supervised learning and image fusion. Neural Process. Lett. **53**(2), 1253–1285 (2021)
17. Barbedo, J.G.A., Koenigkan, L.V., Santos, T.T., Santos, P.M.: A study on the detection of cattle in UAV images using deep learning. Sensors **19**(24), 5436 (2019)
18. Shafiee, M.J., Chywl, B., Li, F., Wong, A.: Fast YOLO: a fast you only look once system for real-time embedded object detection in video. arXiv preprint arXiv:1709.05943 (2017)
19. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M: YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)

20. Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H.: Attention branch network: learning of attention mechanism for visual explanation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10 705–10 714 (2019)
21. Li, X., Jia, X., Wang, Y., Yang, S., Zhao, H., Lee, J.: Industrial remaining useful life prediction by partial observation using deep learning with supervised attention. IEEE/ASME Trans. Mechatron. **25**(5), 2241–2251 (2020)
22. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018)
23. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
24. Hu, J., Zhi, X., Shi, T., Zhang, W., Cui, Y., Zhao, S.: Pag-YOLO: a portable attention-guided yolo network for small ship detection. Remote Sens. **13**(16), 3059 (2021)
25. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
26. Wang, C., Luo, Q., Chen, X., Yi, B., Wang, H.: Citrus recognition based on YOLOv4 neural network. In: Journal of Physics: Conference Series, vol. 1820, no. 1, p. 012163. IOP Publishing (2021)
27. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28, pp. 91–99 (2015)
28. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
29. Kaixuan, Z., Dongjian, H.: Target detection method for moving cows based on background subtraction. Int. J. Agric. Biol. Eng. **8**(1), 42–49 (2015)