ICIC Express Letters
Part B: Applications
Volume **13**, Number **12**, December **2022**

ICIC International ©2022 ISSN 2185-2766
pp. 1313–1319

BLACK COW TRACKING BY USING DEEP LEARNING-BASED ALGORITHMS

Cho Cho Aye[1], Thi Thi Zin[1,*] and Ikuo Kobayashi[1,2]

[1]Graduate School of Engineering
[2]Sumiyoshi Livestock Science Station, Field Science Center, Faculty of Agriculture
University of Miyazaki
1-1, Gakuen Kibanadai-Nishi, Miyazaki 889-2192, Japan
z321702@student.miyazaki-u.ac.jp; ikuokob@cc.miyazaki-u.ac.jp
*Corresponding author: thithi@cc.miyazaki-u.ac.jp

Received April 2022; accepted June 2022

Abstract. *Raising livestock is essential in the farming industry to meet the consumer's requirement. Livestock monitoring system is useful to monitor their health without needing too much manpower. Thus, livestock tracking becomes one of the vital parts of livestock monitoring system. The objective of this proposed system is to track black cows based on detected features. Here, the YOLOv5 (You Only Look Once) model was used in detection phase to detect cow regions and Deep SORT (Simple Online Real-time Tracking) was applied to tracking the target cows in every consecutive frame. In Deep SORT, it includes appearance feature model to recognize cow's visual appearance such as shape, size, and pose. The proposed system was best trained by adopting transfer learning method. The detection model achieves an accuracy of 0.995 mAP@0.5 whereas the tracking model gets the performance results in video-1 and video-2 with 99.4% and 98.9%, respectively.*
Keywords: YOLOv5, Deep SORT, Cow detection and tracking, Transfer learning, Black cow

1. **Introduction.** With the world population growth, the demand for the food supply has increased significantly. The livestock management system takes a vital role in monitoring cow activities and actions to ensure their health condition. Animal behaviour analysis such as detection of abnormal behaviour, detecting estrus and lameness was proposed by using image processing techniques [1]. Various studies also introduced detection of mounting behaviours in cattle by using deep learning-based algorithms for detecting the estrus state which is also one of the important parts of livestock monitoring system [2,3]. Extracting the region of cow with the help of image processing technology was also studied by authors in [4]. This extracted cow region is essential for further cow behaviour analysis. For black cows in this proposed system, they are mainly raised for meat production so that they are mostly kept in the pens and fed for weight gain as much as possible without too much physical movement. As a result, if the cow is suffered from serious illness that cannot be found out in time, it can cause cow fatal. Therefore, the ranchers can monitor the abnormal behaviour and daily movement of the individual cow based on this proposed tracking system by reducing the cow fatal rate and increasing the revenue.

Although there are many livestock tracking systems that were contributed by other researchers [5,6], the wearable sensors or smart tags were used, and they gave burden to cows, and it is expensive to mount in every cow. With the help of computer vision techniques, tracking livestock and people can be performed by using cameras instead of installed sensors to the target objects. Pig tracking system with Faster R-CNN (Region Based Convolutional Neural Networks) and Deep SORT (Simple Online Real-time Tracking) was proposed by researchers in [7] by using the image data from camera with deep

DOI: 10.24507/icicelb.13.12.1313

learning techniques. Moreover, multiple people tracking was also implemented by other researcher by using YOLOv3 (You Only Look Once) and Deep SORT (Simple Online Real-time Tracking) [8]. Also, the authors in [9] also performed object tracking with the help of Deep SORT and YOLOv3 architecture. The tracking result of these methods got high accuracy.

However, people tracking will be less challenging than livestock tracking because the texture features such as cloth colours, hair styles and shapes of each person are quite different. As a result, it is easy to recognize the individual people features. Diary cow detection and tracking method was also introduced with a combination of image processing techniques and deep learning concepts and dairy cows have different skin patterns which are the useful information for more easily performing tracking [10]. In this work, black cow tracking was implemented by analyzing and identifying from the studies above. In addition, it is more challenging tasks because of their same or similar skin patterns. The YOLOv5 model architecture was adapted to the proposed system using transfer learning to detect the location of black cows in each frame of an RGB video input. These detected features were then used to process the tracking algorithm to generate the ID number of each cow. The details of data preparation and the two methods are described in Section 2. Section 3 explains the experimental results and Section 4 is about the conclusion of this proposed system.

2. **Proposed Methodology.** This section mainly includes the process of data preparation for training the required model, and the methods which were used to detect the cow region and to track the individual cows. The process of cow tracking system is emphasized on cow detection and then followed by cow tracking. The overview of the system is shown in Figure 1. At first, the cow region detection was performed on each frame in the video through the optimized YOLOv5 algorithm. Based on the detection result from the YOLOv5, the Deep SORT algorithm was utilized to track the individual black cow.

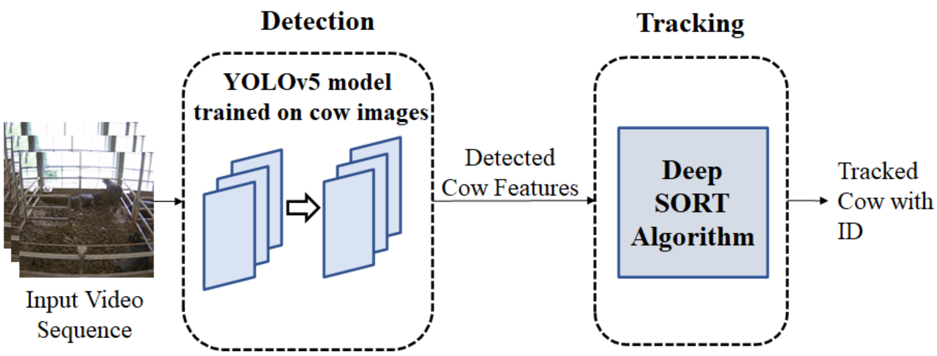

FIGURE 1. The overview architecture of the cow tracking system

2.1. **Dataset preparation.** There were two groups of datasets: 1) cow detection dataset for training YOLOv5 model, and 2) cow Re-ID (Re-identification) dataset for training deep association metric in Deep SORT. The images used in these cow datasets were collected from Sumiyoshi Livestock Science Station, University of Miyazaki, Japan. Each image consisted of six cows from two pens. For detection, the dataset was manually annotated in YOLO format with the class label, centre coordinates (x, y), width and height information. Moreover, the black cow dataset consisted of total 2,600 images and it was randomly divided into two groups: 70% of the black cow dataset was used for the training model and 30% of the black cow dataset for the validation dataset. Concerning the cow Re-ID dataset, it was gathered based on the MARS which is the person Re-ID dataset [11]. This cow Re-ID dataset mainly consisted of 6 cow identities, each identity has 200 images, a total of 1,200 images. Each cow was cropped from annotated bounding box region of the detection dataset. For the training of deep association metric, the cow Re-ID dataset was randomly split into 80% for the training set and 20% for the validation set.

2.2. **Detection method.** In order to achieve good cow tracking results, it is necessary to accurately and quickly detect the target first. The detection method using deep learning is particularly prominent in accuracy and speed performance. Although there are many YOLO object detection models available such as YOLOv3 [12] and YOLOv4 [13], the YOLOv5 model was chosen because of its outperformance in comparison with the previous YOLO models. The YOLOv5 is a one-stage detector that has three main important parts [14]: 1) Backbone: CSP-Cross Stage Partial Network, 2) Neck: PANet, and 3) Head. In YOLOv5, CSP is used as a backbone to extract important features from the input images. PANet is applied to getting feature pyramids to identify the same object with different scaling. The model head is mainly utilized for performing the final detection part by applying anchor boxes on the features and generating output vectors with class probabilities, object scores and bounding boxes information. YOLOv5 has 4 types of models: YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x starting from smallest to largest model. In this system, the YOLOv5x model was chosen to get higher accuracy.

To develop cow detection model using YOLOv5, the network-based transfer learning method was adopted. According to a survey of deep transfer learning, network-based deep transfer learning means the reuse of a partial network with its network structure and connection parameters pre-trained in one model to another model with a specific class [15]. By using the transfer learning method, it can not only save resources but also improve efficiency when training new models. The overall framework of cow detection using YOLOv5 with transfer learning is represented in Figure 2. The model was mainly trained on pre-trained COCO weights with the help of transfer learning. Here, the prepared training images were passed through the YOLOv5x model to train the model for recognizing the black cows. The parameters used for training YOLOv5 are image size: 416, training epoch: 100, learning rate: 0.01, and batch size: 32. After getting the new learned features for cow detection, the test images can be applied and each black cow will be detected with bounding boxes.

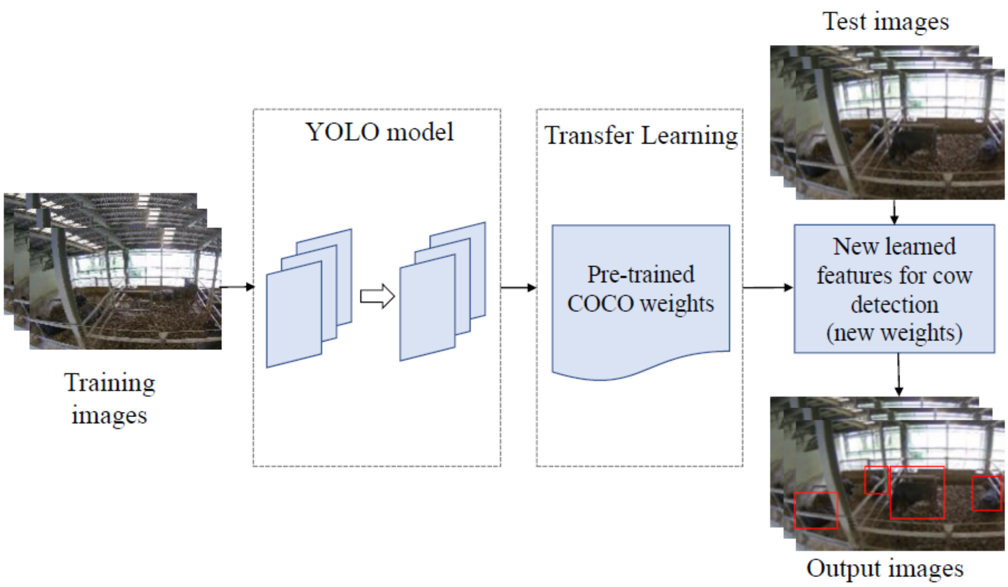

FIGURE 2. Overall framework of cow detection using YOLOv5 with transfer learning

2.3. **Tracking method.** When the trained YOLOv5 model can detect the cow region in each frame, it was required to track the identity of each cow within frames. To achieve this, the Deep SORT algorithm was employed. Deep SORT is the distance and visual-based tracking with a combination of Kalman filter and Hungarian algorithm [16]. The Kalman filter is mainly used to estimate or update the next position of the target object based on the association result. It is represented by an 8-dimensional vector (x, y, a, h, vx, vy, va, vh), where "x, y" is the bounding box centre-coordinates, "a" is the aspect ratio, "h" is the height, and the rest "vx, vy, va, vh" are the respective velocities in image coordinates. On the other hand, the Hungarian algorithm is used to match target tracks with the detections in the current frame. In Deep SORT, the Hungarian algorithm includes not only motion metric (Mahalanobis distance) but also appearance features (cosine distance) which was trained on the CNN (Convolutional Neural Network) model by using cow Re-ID dataset. The general framework of cow tracking with Deep SORT algorithm can be seen in Figure 3. Firstly, the detected cow features from YOLOv5 will be associated with current tracked IDs by using Hungarian algorithm. Based on the association result, the Kalman filter will determine whether to assign the previous same ID or initialize new track with new ID.

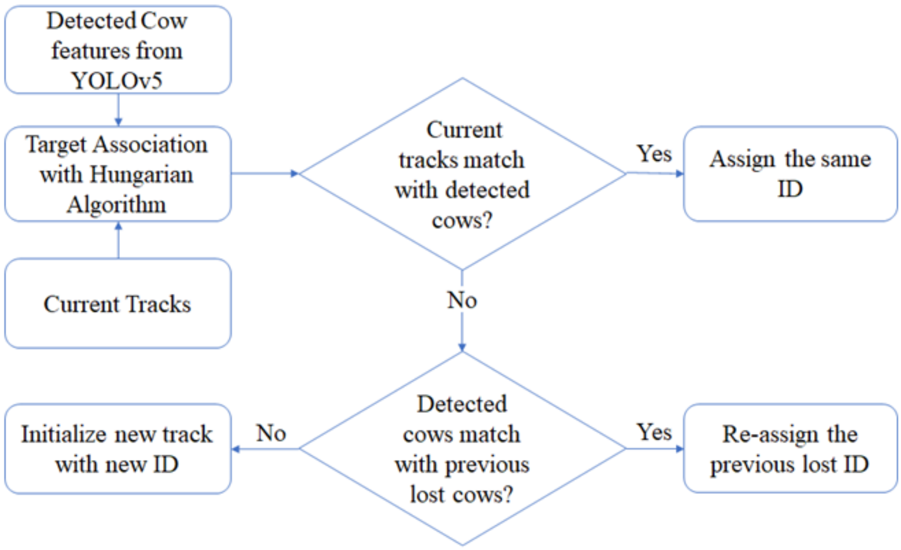

FIGURE 3. General framework of cow tracking using Deep SORT algorithm

3. **Experimental Results.** In this section, the detection results of the YOLOv5 model and the tracking accuracy of Deep SORT from specific testing video results will be presented.

3.1. **Detection result.** To get the optimal detection accuracy, the YOLOv5 model was trained with different epochs such as 100, 200, and 300. In this cow detection system, the training model with the 100 epochs got better performance than the others according to the testing videos result. Moreover, the precision and recall value of the trained model was also computed. The equations used to calculate the precision, recall, and accuracy are shown as

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

The TP (True Positive) in the above equation represents the bounding boxes with the target object were correctly detected and FN (False Negative) means the exiting target object was not detected. FP (False Positive) is represented when the background was wrongly detected as a cow. TN (True Negative) indicates probability of negative class in image classification. In this study, it is ignored in the evaluation process. For the evaluation metrics, the performance of the detection model got 0.995 mAP@0.5. The detection result of the back cow can be seen in Figure 4 where each cow was correctly detected with the cow label and confidence value in the top-left corner of each bounding box.

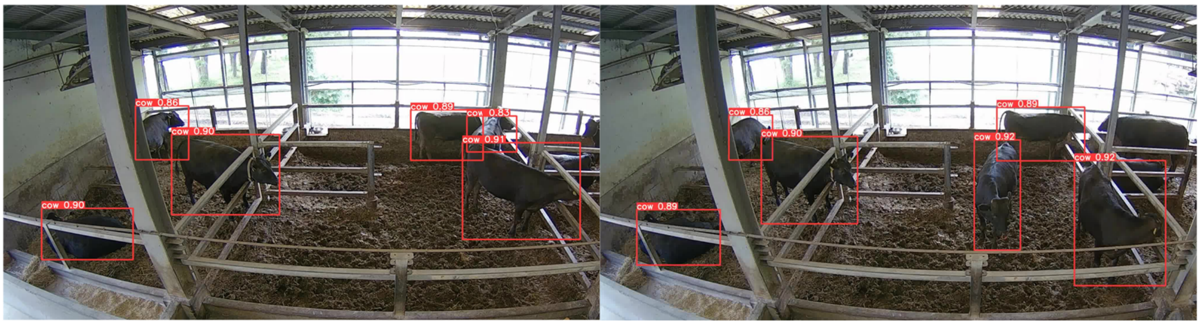

FIGURE 4. Black cow detection with YOLOv5

3.2. **Tracking result.** After applying Deep SORT to the detection result, the black cow can be tracked with a series of ID numbers. Here, the tacking is performed on six black cows in two pens. Thus, the initial ID of these cows started from 1 to 6. To evaluate the tracking accuracy, the most widely used metric MOTA [17] (Multiple Object Tracking Accuracy) will be calculated by deciding when the tracker wrongly detects the other object instead of the target object FP, when the tracker cannot detect the exiting cow in a frame FN when the tracks switch their identities, Identification Switch *(IDSW)*. The equation of MOTA is

$$\text{MOTA} = 1 - \frac{\sum_t FN_t + FP_t + IDSW_t}{\sum_t GT_t} \tag{4}$$

The accuracy of Deep SORT was tested on two videos and each video is 5 minutes long with total of 300 extracted frames. In video-1, the model achieves an accuracy of 99.4% MOTA. In total, there were 1 FPs, 5 FNs, 3 $IDSW$s based on 300 frames. In video-2, the model got an accuracy of 98.9% MOTA with 13 FPs, 6 $IDSW$s. Some testing results of black cow tracking are shown in Figure 5 where the tracking IDs starting from 1 to 6 were generated and they were applied to specific cow during the whole 5 minutes long videos.

FIGURE 5. Black cow tracking result with ID numbers

4. **Conclusion.** This proposed system introduced the fusion of YOLOv5 and Deep SORT for tracking individual cows which are kept in groups. With the help of transfer learning, the YOLOv5 model for black cow detection was implemented with better accuracy and a faster training time. For black cow tracking, the detection model was combined with the Deep SORT tracking algorithm. Then, the output features from the YOLOv5 model were used as input features of Deep SORT. After matching the target and the detection as an association problem, the Deep SORT produced the related tracking result with ID numbers. Moreover, the detection phase got a 0.995 mAP@0.5 and the tracking result achieved 98.9% or 99.4% of MOTA without using additional hardware or visual aids.

However, tracking the black cow is a challenging work because all the cows have the same or similar textures features which are not easy to recognize even with human eyes. Therefore, the proposed system remains some ID switching and false detection problems to be solved. In future work, the model will be improved with more datasets for reducing the detection problem. In addition, the long-time occlusion problem still needs to be solved because the black cow which was kept in the pen is less active and the occlusion chance is longer than pedestrians.

REFERENCES

[1] T. T. Zin, I. Kobayashi, P. Tin and H. Hama, A general video surveillance framework for animal behavior analysis, *Proc. of the 3rd Intl. Conf. on Computing Measurement Control and Sensor Network (CMCSN-2016)*, Matsue, Japan, pp.130-133, 2016.

[2] S. M. Noe, T. T. Zin, P. Tin and I. Kobayashi, Automatic detection of mounting behavior in cattle using semantic segmentation and classification, *Proc. of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech2021)*, Nara, Japan, pp.227-228, 2021.

[3] S. M. Noe, T. T. Zin, P. Tin and I. Kobayashi, Automatic detection and tracking of mounting behavior in cattle using a deep learning-based instance segmentation model, *International Journal of Innovative Computing, Information and Control*, vol.18, no.1, pp.211-220, 2022.

[4] Y. Motomura, T. T. Zin and Y. Horii, Cattle region extraction using image processing technology, *Proc. of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE2021)*, Kyoto, Japan, pp.762-763, 2021.

[5] S. Koompairojn, C. Puitrakul, N. Riyagoon and S. Ruengittinun, Smart tag tracking for livestock farming, *The 10th International Conference on Ubi-Media Computing and Workshops (Ubi-Media)*, pp.1-4, 2017.

[6] A. Muminov, D. Na, C. Lee and H. S. Jeon, Virtual fences for controlling livestock using satellite-tracking and warning signals, *International Conference on Information Science and Communications Technologies (ICISCT)*, pp.1-7, 2016.

[7] J. Cowton, I. Kyriazakis and J. Bacardit, Automated individual pig localisation, tracking and behaviour metric extraction using deep learning, *IEEE Access*, vol.7, pp.108049-108060, 2019.

[8] I. Ahmed, M. Ahmad, A. Ahmad and G. Jeon, Top view multiple people tracking by detection using Deep SORT and YOLOv3 with transfer learning: Within 5G infrastructure, *International Journal of Machine Learning and Cybernetics*, vol.12, no.11, pp.3053-3067, 2021.

[9] T. L. Dang, G. T. Nguyen and T. Cao, Object tracking using improved Deep_Sort_YOLOv3 architecture, *ICIC Express Letters*, vol.14, no.10, pp.961-969, 2020.

[10] C. C. Mar, T. T. Zin, I. Kobayashi and Y. Horii, A hybrid approach: Image processing techniques and deep learning method for cow detection and tracking system, *Proc. of the 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech2022)*, Osaka, Japan, pp.566-567, 2022.

[11] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang and Q. Tian, MARS: A video benchmark for large-scale person re-identification, *European Conference on Computer Vision*, pp.868-884, 2016.

[12] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *arXiv Preprint*, arXiv: 1804.027 67, 2018.

[13] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv Preprint*, arXiv: 2004.10934, 2020.

[14] J. Ieamsaard, S. N. Charoensook and S. Yammen, Deep learning-based face mask detection using YOLOv5, *The 9th International Electrical Engineering Congress (iEECON)*, pp.428-431, 2021.

[15] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, A survey on deep transfer learning, *International Conference on Artificial Neural Networks*, pp.270-279, 2018.

[16] N. Wojke, A. Bewley and D. Paulus, Simple online and realtime tracking with a deep association metric, *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, pp.3645-3649, 2017.
[17] K. Bernardin and R. Stiefelhagen, Evaluating multiple object tracking performance: The CLEAR MOT metrics, *EURASIP Journal of Image Video Processing*, pp.1-10, 2008.