International Journal of Innovative Computing, Information and Control
Volume **18**, Number **1**, February **2022**

ICIC International ©2022 ISSN 1349-4198
pp. **211–220**

AUTOMATIC DETECTION AND TRACKING OF MOUNTING BEHAVIOR IN CATTLE USING A DEEP LEARNING-BASED INSTANCE SEGMENTATION MODEL

Su Myat Noe[1,*], Thi Thi Zin[1], Pyke Tin[1] and Ikuo Kobayashi[2]

[1]Graduate School of Engineering
[2]Field Science Center, Faculty of Agriculture
University of Miyazaki
1-1, Gakuenkibanadai-Nishi, Miyazaki 889-2192, Japan
{ thithi; ikukob }@cc.miyazaki-u.ac.jp; pyketin11@gmail.com
*Corresponding author: z319701@student.miyazaki.ac.jp

Received July 2021; revised October 2021

Abstract. *In precision livestock farming, estrus detection in cattle is particularly important for cattle breeding management. With accurate estrus detection, artificial insemination can be administered, which proportionally affects the productivity of livestock farms. Most estrus behaviors can be successfully detected by recognizing the mating postures of cattle. Therefore, in this paper, we propose an estrus detection approach that tracks and identifies cattle mating postures individually based on video inputs. To achieve precise identification and to obtain individual cattle information, segmenting each cattle from its background is a vital step. To solve pixel-level segmentation masks for the cattle in an outer ranch environment, an instance segmentation approach based on a Mask R-CNN deep learning framework is also proposed. In this paper, individual cattle segmentation for detecting the mounting behaviors is carried out first. This is followed by a lightweight tracking algorithm as a post-processing step which is our study innovation. The training data were collected by installing surveillance cameras at a livestock farm, and for the testing data, various datasets from different camera placements were used. The proposed approach achieved 95.5% detection accuracy in identifying the estrus behaviors of cattle.*
Keywords: Cattle mounting behavior, Cattle tracking, Deep learning, Mask R-CNN

1. **Introduction.** The livestock industry has developed noticeably in recent years and cattle breeding management has become an essential factor in the developing industry. Among other things, estrus detection is particularly important for cattle breeding management. After estrus detection, cattle are artificially inseminated by livestock farm managers. If estrus behaviors are not detected timely, the short window for artificial fertilization may be missed and the infertility period of cattle may persist [1,2]. Thus, a successful breeding management program is necessary for livestock farms which in turn emphasize the importance of estrus detection. Traditional monitoring methods such as tail paint rely on human observations which is time consuming and requires a lot of effort. It is also prone to confusion at times.

With the advancement in the field of Artificial Intelligence (AI), numerous studies on cattle estrus detection using AI technologies have been conducted. Most estrus detections use sensors such as pedometers and accelerometers that are often attached to cattle's hind limbs to capture the increment in the number of steps during cattle estrus behaviors. Those detection methods also consist of an acceleration sensor and a gyro sensor that

DOI: 10.24507/ijicic.18.01.211

measure the cattle momentum as well as potential energy and notify the manager whenever an estrus behavior occurs [3,4]. Although these techniques achieve high accuracy, they also introduce problems such as causing stress to cattle and increases in the overall cost due to the equipment. Moreover, the sensors may record false readings and require direct contact with cattle which affects animal welfare. Other scholars also experimented Infrared Thermography (IRT) in detecting cattle estrus behaviors [5,6].

To solve the above problems, this study proposes the development of non-contact video monitoring systems that would help in reducing stress for both cattle and livestock farmers. Although a few studies on the detection of cattle estrus behaviors using the camera are deployed into the real world, the end results can be varied when the environment changes since they are using traditional image processing methods [7]. It is critical to effectively detect mounting behaviors and further monitor the cattle estrus behaviors in the complex natural environment. Therefore, a data-driven estrus detection would be the next step that could solve those issues.

In this research, we proposed building an estrus detection system based on video inputs of cattle by using a deep learning approach. Estrus detection is achieved through the identification of cattle mating posture, which is a specific behavior expressed during estrus [8,9]. To automatically detect the cattle behaviors, the most challenging part is to detect the foreground object, the cattle, from the video inputs. The illumination conditions, cluttered backgrounds, and occlusion of the object impose the challenge to recognize the cattle in an outdoor ranch.

With the help of deep learning, computer vision techniques such as object detection and semantic segmentation have drastically advanced [10]. In this paper, we applied object detection, a method of deep learning-based image analysis. We leveraged base object detection algorithms by using a customized Mask R-CNN networks with additional layers to improve the detection performance in various environments. The novelty of our system is as a post-processing method, we combined tracking of each cattle by implementing our lightweight tracking algorithm based on the Mask R-CNN output results. Although there is no current research on Mask R-CNN and customized lightweight tracking, according to the literature review, in [3]. They have received 92% accuracy compared to our results of 95.5% accuracy. So ours is 3.5% better accuracy. The findings of this research will promote successful breeding management for both small and large farms in the cattle industry. Such advantages consequently contribute to the economic growth of the farms.

2. **Overall Proposed System.** The details of the workflow procedure are proposed in this section. For individual cattle monitoring and their welfare measurement in vision-based precision livestock farming, effective cattle segmentation is necessary for further image analysis. The Mask R-CNN method was proposed to recognize cattle instance segmentation in the complex feedlot environment as presented in Figure 1. For the object detection (Instance Segmentation) part, we used the Mask R-CNN method. A data acquisition is explained in Section 2.1. Moreover, a detailed detection by Mask R-CNN is proposed in Section 2.2. The automatic feature extraction and lightweight tracking algorithm are described in Sections 3 and 4.

2.1. **Data acquisition.** In this study, Japanese black cattle in the Sumiyoshi Livestock Science Ranch were our research targets. Cattle were free to move because they are kept in large outdoor ranches. They are walking and laying freely on the ground in the open-air ranch. The cattle were continuously captured by a GV-FER5700 camera located at the top of the barn providing the best possible view of the whole barn. The camera has a resolution of 2560×2048 pixels, recording at 30 frames per second.

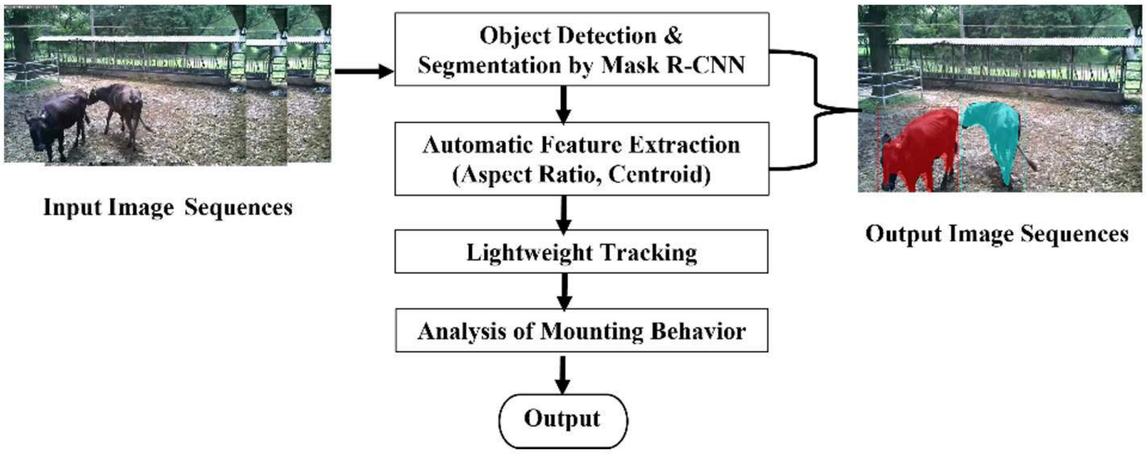


FIGURE 1. The major workflow of our proposed system

In our project, Japanese black cattle data were collected in 24 hours. Image sequences were acquired when the cattle were walking on the ranch. This cattle dataset is challenging to process due to the following aspects: 1) the cattle change posture frequently; 2) high comparison between cattle from the same types due to the similar coat color; 3) lighting is changing. Images acquired in the morning are low light, while they are over illuminated and have strong shadows at noon. These lighting problems can make machine algorithms inaccurately learn these patches or shadows as cattle's features; 4) the complex background (including tree and ground) is difficult to be distinguished from the segmented cattle. In some rare cases, the wet soil background due to the rain cannot be differentiated from the cattle.

2.2. **Detection by Mask R-CNN network.** The Mask R-CNN model extends Faster R-CNN model by including instance segmentation in addition to object detection and classification. The model coincides with Faster R-CNN about the backbone of Convolutional Neural Network (CNN) and the Region Proposal Network (RPN). The backbone of the Mask R-CNN implementation used ResNet50 or ResNet101 (Residual Network) with an additional Feature Pyramid Network (FPN). In this study, the powerful ResNet101 was selected as the backbone, which is shown in Figure 2. The reason for choosing ResNet101 as a backbone network is that it has a score of 98.3% accuracy in animal farming and

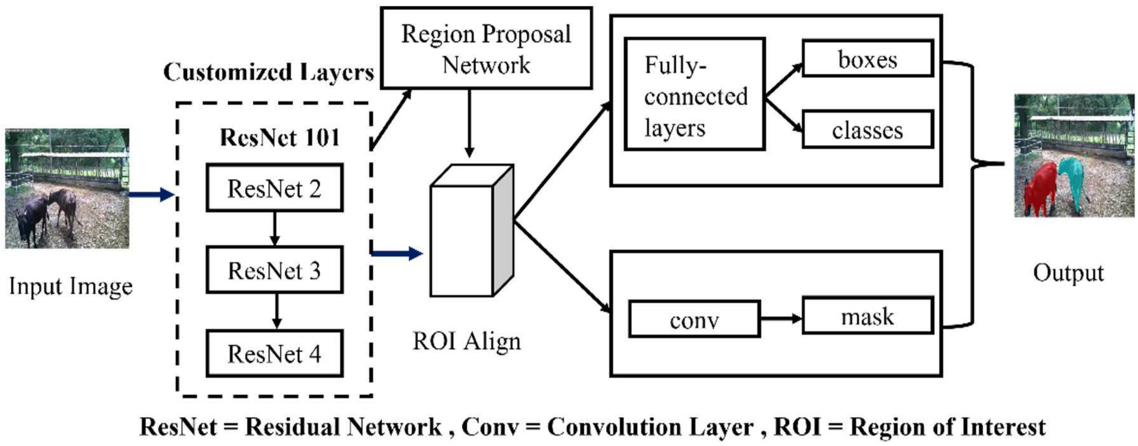


FIGURE 2. Framework of Mask R-CNN network architecture on cattle segmentation

78.3% accuracy in ImageNet challenge. The backbone network ResNet101 extracts features from the input image, starting with low-level features like edges and corners specified by early layers, and in the later layers.

2.3. **Implementation and training of Mask R-CNN.** The proposed cattle instance segmentation model was based on the Mask R-CNN implementation by Matterport using the Keras application programming interface. The requirements are Python 3.8.11, TensorFlow 2.5.0, and Keras 2.0.8 with GPU. The base model was pre-trained on the MS COCO dataset and transfer learning approach was used for the proposed model. The implementation was performed on the hardware of GPU model of GeForce GTX 1080 Ti and 1 TB Hard disk.

2.4. **Learning rate scheduler and data augmentation.** The Mask R-CNN model was enhanced manually with a regulated reduction of the learning rate and data augmentation applied to the training data. Decreasing the learning rate during training is based on the idea that a high learning rate enables a faster approach to the minimum of the loss function. A learning rate scheduler was defined starting with a 0.01 initial learning rate. The scheduler was added to the Keras callbacks. In the data augmentation step, the current batch of images was run through a pipeline of transformations before inputting into the model for training. A Python module 'imgaug' was loaded and four augmentation techniques were applied. All augmenters were applied to a random 50% of the images in a batch and carried out in random order: 1) Flip Augmentation; 2) Brightness Augmentation; 3) Gaussian Blur [11,12]. For solving the problem of difficult training of a small-scale training dataset (∼1600 images), the transfer learning method was utilized to train the cattle segmentation network [13-16].

2.5. **Training.** As the model was already pre-trained, here only stages four and five layers of the ResNet101, FPN backbone, and all rest of the parts of the model were trained. The training was run for 30 epochs. The whole training time was around 1 hour for each network.

2.6. **Training and validation datasets.** After analyzing a total of 2000 images, 1600 images were used for training and 400 images for validation respectively. The images are of 1024×768 dimensions. To make the sample flexible and contain cattle daily posture, images that include walking, standing, lying, and mounting behaviors of cattle in different illumination and cleanliness conditions were selected. These images were drawn from the videos and manually annotated by using the free online tool VGG Image Annotator (VIA 3). The detailed dataset production process is described in Figure 3.

2.7. **Evaluating the model.** To evaluate the performance of cattle segmentation and contour extraction, Mean Pixel Accuracy (MPA) is used, respectively.

$$MPA = \frac{1}{k} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{1}$$

where k is the total number of categories including the background, p_{ij} is the total number of pixels whose real pixel class is i and predicted as j, and p_{ii} is the total number of pixels whose real pixel class is i but predicted as i.

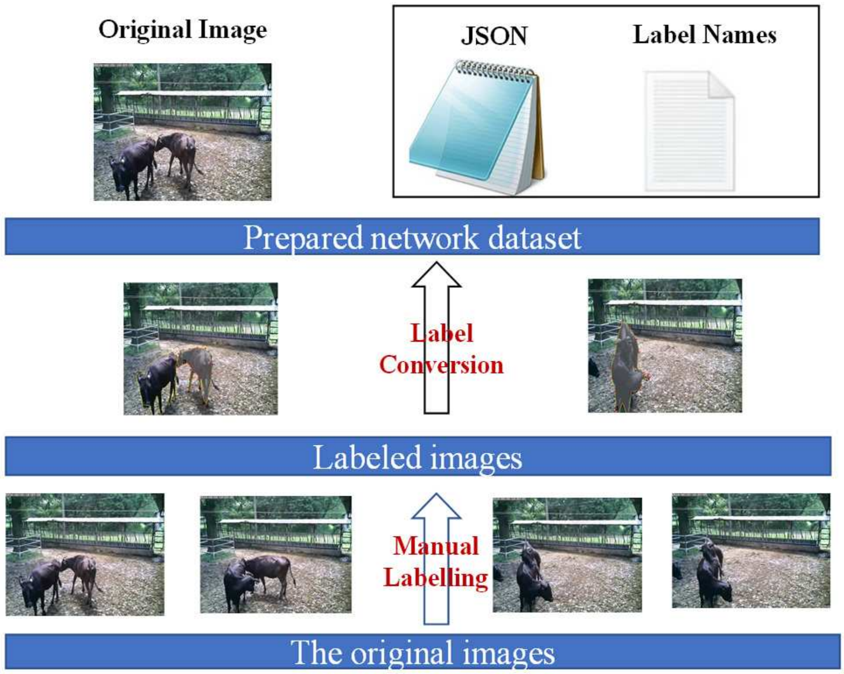


FIGURE 3. Dataset production process

FIGURE 4. Cattle segmentation using the proposed Mask R-CNN

2.8. **Segmentation results.** The segmentation results on the testing image datasets results of the proposed Mask R-CNN approach can be seen in Figure 4. If two cattle in the mounting behavior occur in the ranch, the proposed approach can still detect and segment each cattle with different colors (i.e., green, blue, or red). However, the Mask R-CNN effectively visualized "cattle" objects with confidence scores of mAPs 95.5% and F1 scores 75.6% as shown in Figure 4. These results suggest additional training time, dataset acquisition, and data cleaning to attain higher prediction scores for both models. The cattle segmentation accuracies are shown on the output cattle image.

3. **Feature Extraction.** Based on the accurate detection of cattle regions, the efficient feature extraction of regions was crucial to identify and track each cattle. There are two obvious differences when comparing the mounting behavior with the other behaviors: the mounting posture and the quick upward movement of the cattle's body. To reduce the complexity of feature extraction, subsequent processing was based on the Mask R-CNN output of the detected regions. The geometric features of detected regions were extracted by processing the difference between corresponding regions in consecutive frames, which were extracted through the calculation. After detecting the cattle bounding box with Mask R-CNN, the information is further processed by the lightweight tracking algorithm. Feature extraction was performed by applying the 5 features extracted by each of the cattle bounding box objects, width (w), height (h), aspect ratio (r), centroid (x, y). Firstly, the position p at time t of the detected cattle object is defined in Equation (2).

$$p(t) = (x(t), y(t)) \tag{2}$$

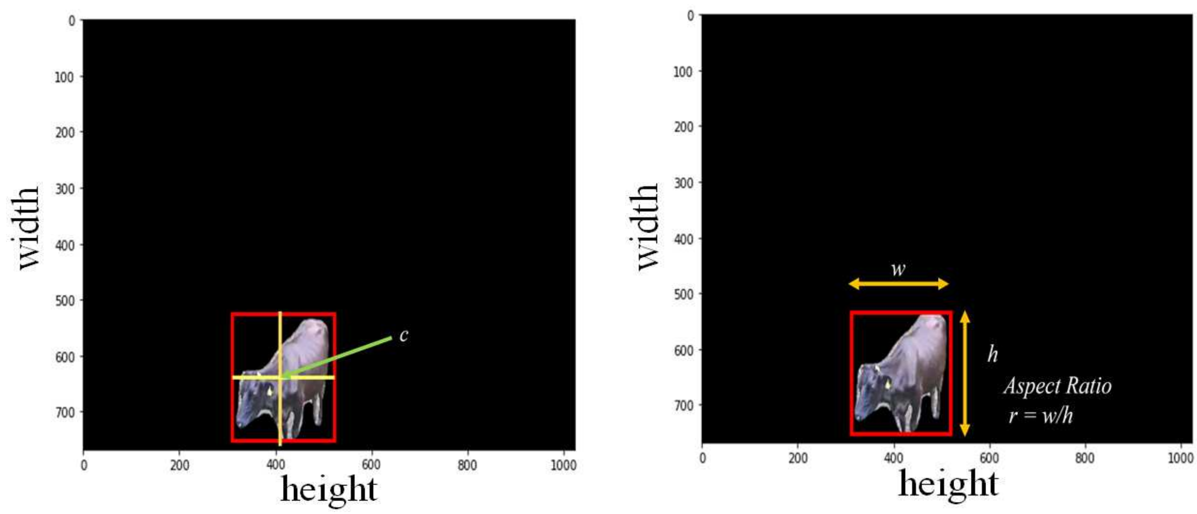
Secondly, the centroid of the object is obtained by Equation (3),

$$C(t) = (x_c(t), y_c(t)) \tag{3}$$

Finally, the aspect ratio (r) concerning the object was calculated to estimate the cattle posture as shown in Equation (4).

$$r(t) = w(t)/h(t) \tag{4}$$

where $r(t)$ is the aspect ratio of the object found by dividing the value of width (w) by that of the height (h) for the object at time t. The concepts for deriving the aspect ratio were illustrated in Figure 5.



(a) (b)

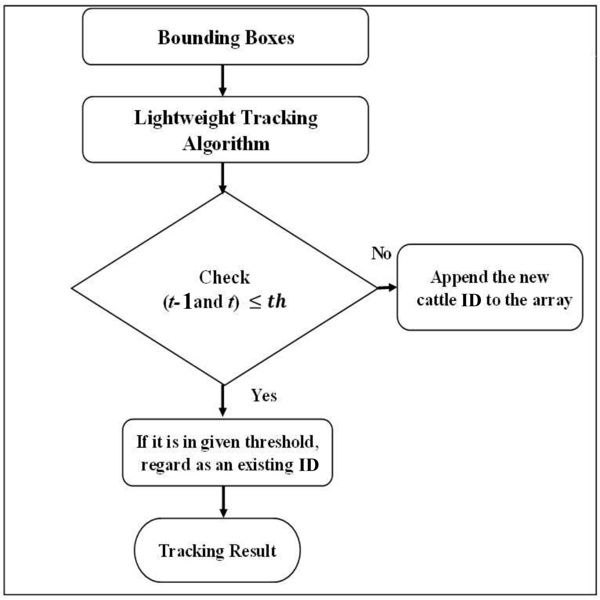
Figure 5. Feature extraction: (a) Centroid (c); (b) aspect ratio of the object (r)

4. **Lightweight Tracking Algorithm.** Once the feature extraction was completed, the tracking was performed. Our current proposed tracking algorithm covers the function of Kalman filter and Hungarian algorithm, which enables to recover tracks between missed detections and less ID switch cases while tracking was performed. The tracking of a sequence of each frame in the bounding box was defined by the following Equations (5) and (6).

$$|BB_{t-1}(F_i) - BB_t(F_i)| \leq Th \quad \text{for } i = 1, \ldots, 6 \tag{5}$$

$$F = [x, y, w, h, gc_x, gc_y] \tag{6}$$

where BB is the bounding box, $t-1$ is the previous frame, Th is the threshold, F is the total number of features, i is the six features, x, y, w, h are the width and height value of the bounding box, gc_x and gc_y are the centroid x and y values. In Figure 6, the 5th line, the max threshold refers to the process of adding the new cattle ID to the array and min threshold refers to the definition of the existing ID. The max and min values are got by doing the trial-and-error method. Moreover, $o_{i,j}$ refers to the overlap of two instances i and j. Our proposed tracking algorithm as a post-processing method is based on bounding box values as we extract from Mask R-CNN instance segmentation method. The label number in Figure 7 refers to the tracking ID of each cattle, which results output from our lightweight tracking algorithm. Less ID switching during tracking time is the superiority of our research. Thus, the farm manager can accurately decide which cattle show mounting events by only looking at the cattle ID number which were described in the Figure 7. Also, it helps to decide the optimal time to the artificial insemination. Each consecutive image x has each instance i. Instances are matched according to the geometric centers that are close. For this purpose, we track instances according to their next frame using information from previous frames. We denote by $\{x_t\colon t = 1, \ldots, T\}$ the image sequence with a length of T. Our tracking algorithm was applied sequentially to each frame $t = 1, 2, 3, \ldots, T$.



Algorithm: Lightweight Tracking Algorithm

/* **shift** */
if *t and instance i exists in frame t*-1 **then**
    shift instance i from frame t-1 by the vector $(\bar{i}_t - \bar{i}_{t-1})$
    if $\max_{j \varepsilon \bar{I}x_t} o_{i,j} \geq threshold$ **then**
        match instances i and j
    else if $\min_{j \varepsilon \bar{I}x_t} o_{i,j} \leq threshold$ **then**
        match instances i and j
/* **lost and enter new case***/
if t-1 and i does not exist in frame t **then**
    search the value in t-2:
    if *(t-2)* contains the value of last instance **then**
        match instances i and j
        stop the matching process
    else
        do not match instances i and j
        define the new instance ID and append it

FIGURE 6. The proposed lightweight tracking algorithm

5. **Conclusions.** In this study, we proposed the identification of the mating posture of cattle for estrus detection. Deep learning-based object detection was used to identify the mating postures of cattle which is a robust cattle segmentation Mask R-CNN instance segmentation of each cattle. As a post-processing method, a lightweight tracking algorithm was proposed. Mask R-CNN was used to identify cattle regions in a complex ranch environment. Our experimental results demonstrated that our identification system for mounting action detection can perform well in a condition where a small amount of training data ($\sim$1600 images) is given. The detection accuracy rates were 95.5% which was a satisfactory result, measured by mAPs metric. In future research, we intend to

FIGURE 7. Our proposed lightweight tracking algorithm results

further improve the segmentation results and we will carry out analysis of mounting behavior as it needs a lot of work to be done. Based on these extracted features, analysis of decision-making rules of prediction mounting behaviors will be built using a hidden Markov model to figure out the best time for artificial insemination. For the comparison approach, machine learning methods will be applied for time-series data.

Acknowledgment. This work was supported in part by SCOPE: Strategic Information and Communications R&D Promotion Program (Grant No. 172310006) and JSPS KAKENHI (Grant No. 17K08066).

Institutional Review Board Statement. Ethical review and approval were waived for this study, due to no enforced nor uncomfortable restriction to the animals during the study period. The image data in this study were collected by an installed camera without disturbing natural parturient behaviors of animals and routine management of the farm.

REFERENCES

[1] S. M. Noe, T. T. Zin, P. Tin and H. Hama, Detection of estrus in cattle by using image technology and machine learning methods, *Proc. of IEEE 9th Global Conf. on Consumer Electronics (GCCE2020)*, Kobe, Japan, pp.320-321, DOI: 10.1109/GCCE50665.2020.9291987, 2020.

[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 IEEE Global Conference on Life Sciences and Technologies*, Nara, Japan, pp.227-228, DOI: 10.1109/LifeTech52111.2021.9391980, 2021.

[3] N. M. Arago et al., Automated cattle counting using Mask R-CNN in quadcopter vision system, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.11, no.9, pp.303-311, 2020.

[4] A. T. Sarkisov, Beef cattle instance segmentation using fully convolutional neural network, *The 29th British Machine Vision Conference (BMVC)*, 2018.

[5] J. Salau and J. Krieter, Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting, *Animals*, vol.10, no.12, DOI: 10.3390/ani10122402, 2020.

[6] Y. Qiao et al., Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming, *Computer and Electronic in Agriculture*, vol.165, DOI: 10.1016/j.compag.2019.104958, 2019.

[7] Y. Kawano et al., Toward building a data-driven system for detecting mounting actions of black beef cattle, *International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021.

[8] J.-W. Chae and H.-C. Cho, Identifying the mating posture of cattle using deep learning-based object detection with networks of various settings, *Journal of Electrical Engineering & Technology*, vol.16, pp.1685-1692, 2021.
[9] Y. Guo et al., Detection of cow mounting behavior using region geometry and optical flow characteristics, *Computer and Electronic in Agriculture*, vol.163, DOI: 10.1016/j.compag.2019.05.037, 2019.
[10] Y. Qiao et al., Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation, *Computer and Electronic in Agriculture*, vol.185, DOI: 10.1016/j.compag.2021.106143, 2021.
[11] A. Nasirahmadi et al., Deep learning and machine vision approaches for posture detection of individual pigs, *Sensors*, vol.19, no.17, DOI: 10.3390/s19173738, 2019.
[12] D. Li et al., Mounting behavior recognition for pigs based on deep learning, *Sensors*, vol.19, no.22, DOI: 10.3390/s19224924, 2019.
[13] Q. Li, R. Li, K. Ji and W. Dai, Kalman filter and its application, *2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, pp.74-77, DOI: 10.1109/ICINIS.2015.35, 2015.
[14] B. Sahbani, Kalman filter and iterative-hungarian algorithm implementation for low complexity point tracking as part of fast multiple objects tracking system, *IEEE International Conference on System Engineering and Technology*, 2016.
[15] K. Maag, Improving video instance segmentation by light-weight temporal uncertainty estimates, *Computer Vision and Pattern Recognition*, 2021.
[16] Y. Hashimoto, H. Hama and T. T. Zin, Robust tracking of cattle using super pixels and local graph cut for monitoring systems, *International Journal of Innovative Computing, Information and Control*, vol.16, no.4, pp.1469-1475, 2020.

Author Biography

Su Myat Noe received the B.C.Sc. degree (with honor) in Knowledge Engineering in 2018 from University of Information Technology (UIT), in Yangon, Myanmar. Before joining her master's degree, she joined as a research student in University of Miyazaki, Japan. She is currently attending master's program in Energy and Electronics supervised by Dr. Thi Thi Zin, at University of Miyazaki, Japan. Her research interests include cow behavior analysis, image recognition, object tracking, neural network, machine learning, and deep learning. She received the Women in Engineering Award of the international conference LifeTech 2021, which was held physically and virtually on 9-11 March 2021. She is a member of IEEE.

Thi Thi Zin received the B.Sc. degree (with honor) in Mathematics in 1995 from Yangon University, Myanmar and the M.I.Sc. degree in Computational Mathematics in 1999 from University of Computer Studies, Yangon, Myanmar. She received her Master and Ph.D. degrees in Information Engineering from Osaka City University, Osaka, Japan, in 2004 and 2007, respectively. From 2007 to 2009, she was a Post-Doctoral Research Fellow of Japan Society for the Promotion of Science (JSPS). She is currently a Professor of Graduate School of Engineering, University of Miyazaki, Miyazaki, Japan. Her research interests include human behavior understanding, intelligent transportation systems, cow behavior analysis, health care monitoring systems and image recognition. She is a member of IEEE.

Pyke Tin received the B.Sc. degree (with honor) in Mathematics in 1965 from University of Mandalay, Myanmar, the M.Sc. degree in Computational Mathematics in 1970 from University of Rangoon, Myanmar and the Ph.D. degree in stochastic processes and their applications in 1976 from Monash University, Australia. He was the Rector of the University of Computer Studies, Yangon and Professor of Computational Mathematics. He is now a visiting professor of International Relation Center, University of Miyazaki, Miyazaki, Japan. His research interests include image search engines, queueing systems and computer vision, stochastic processes and their applications to image processing.

Ikuo Kobayashi received the D.V.M. degree in Veterinary Medicine from University of Miyazaki, Miyazaki, Japan in 1996. He received his Ph.D. degrees in Veterinary Medicine in 2010 from the United Graduate School of Veterinary Science, Yamaguchi University, Japan. He is currently an Associate Professor of the University of Miyazaki. His research interests include animal science, especially large animal management, reproduction and animal behavior.