ICIC Express Letters
Part B: Applications
Volume **13**, Number **4**, April **2022**

ICIC International ©2022 ISSN 2185-2766
pp. 405–412

A DEEP LEARNING METHOD OF EDGE-BASED COW REGION DETECTION AND MULTIPLE LINEAR CLASSIFICATION

Thi Thi Zin*, Saw Zay Maung Maung and Pyke Tin

Graduate School of Engineering
University of Miyazaki
1-1 Gakuen Kibanadai-nishi, Miyazaki 889-2192, Japan
sawzaymaungmaung@ucsy.edu.mm; pyketin11@gmail.com
*Corresponding author: thithi@cc.miyazaki-u.ac.jp

Received August 2021; accepted October 2021

Abstract. *In this paper we propose a deep learning method of cow region detection and multiple linear model for classifying behaviors of pregnant cows prior to the occurrence of calving events. Dairy farm management experts and farmers have been well recognized that video monitoring to an individual cow plays an important and significant role in production and re-production processes. To be specific we can learn through video monitoring cow's health conditions, body conditions even the occurrences of calving difficulties in times. Moreover, due to the advances in the latest computer vision and image processing algorithms, it is possible to develop a camera system that automatically detects the cow's conditions at a low cost. The fundamental and foremost important step in the proposed system is to detect and segment the cow regions in the video sequences. After the detection process we performed a multiple linear model to classify some behaviors of the detected cows. In particular we consider four states of cow behaviors such as lying state, transition on state from lying to standing, standing state and transition state of standing to lying which are important in studying dairy cow management systems. In order to confirm the validity of our proposed method some experiments are carried out by establishing the video monitoring cameras at the maternity pens of a large dairy farm in Japan. The experimental results show that the proposed method gives an impression of promising with high accuracy.*

Keywords: Deep learning method, Edge-based cow region detection, Holistically-nested algorithm, Multiple linear models, States of cow behaviors, Video monitoring system, Image processing techniques

1. **Introduction.** In optimizing dairy farm management systems, monitoring an individual cow has played an important role in investigating the conditions of cow's health, body conditions, and calving situations. Knowledge of these conditions is important because healthy cows, right body condition scores, and smooth calving events are more productive and leading to smart dairy farming [1-3]. Traditionally, such conditions are done by human experts through visual observations [4]. However, manual observations to analyze individual cows are impractical for commercial farms. Thus, there is increasing demand for automated video monitoring systems in place of human experts. On the other hand, the video monitoring systems have been successfully applied to many areas with security concerns including animal agriculture [5,6]. Moreover, due to the advances in computer vision and image processing techniques, visual animal behavior analysis has become an emerging research field of interest [7-9]. Thus, we can establish a visual monitoring system for dairy cows based on animal appearance and behavior as key indicators of analyzing animal conditions. In this aspect, one of the important steps in spatial locations of cows is to detect and segment the cows within the video sequences. This step can easily be done when a cow walks or stands or is lying individually in a very clear background and

DOI: 10.24507/icicelb.13.04.405

with no obstacles. However, in real-life situations, such ideal environments are very rare and most of video monitoring cameras can capture unnecessary objects such as human interruptions and barn fences making the problem more challenging. Thus, we need a more sophisticated method to detect and segment cow regions.

Many researchers had well recognized that locating cow region only is not sufficient for further statistical analysis. For example, in order to compute the cow body condition scores, assessing an aspect of the animal such as body size or gait, having a binary mask that indicated the cow's location is not enough. The other important information such as joint points or locations of all body parts of a cow is required. This information can then be utilized to interpret in real world terms for human understanding.

In addition, some literature surveys reveal that several papers have reported the problem of detecting edges and object boundaries by using holistically-nested algorithm [9]. It has been long realized that this problem is both fundamental and of great importance to a variety of computer vision areas ranging from traditional image processing tasks to modern applications such as artificial intelligence and Internet of Things environments. Although the holistically-nested edge detection method was applied to many computer visions, to the best of our knowledge, we have not seen any research concerning with dairy cow region detection.

There have been many cow regions segmentation methods focusing on object detection, which generates a binary mask of the objects and their corresponding labels [10-12,17]. Recently, some new methods such as DeepLabCut [13] appeared in the literature to detect animal-related key points. This method requires clear video sequences with a single object and a clear background. However, such requirements are rarely fulfilling in practice. Rather, the real-life videos collected from commercial farms consist of unforeseeable challenges. Accordingly, some issues such as poor illumination [14] and heavy obstacles [15] largely influence the performance of existing detection algorithms.

Apart from stated methods due to the development of convolutional neural networks and deep learning techniques, several approaches for semantic image segmentations have been found in the recent literature. We would like to mention a few of them such as Mask R-CNN [16] and DeepLab [18] which are widely used to analyze the problem of image semantic segmentation.

In summary, we need a video monitoring system that can perform two tasks of segmentation and detection of body motion structure. In addition, how an individual cow behaves or makes movements such as lying, standing, and changes from one posture to another is also important for analysis. Therefore, in this paper, we propose a deep learning method of edge-based cow region detection and multiple linear models for the classification of cow movements. Specifically, we employed a deep learning model of holistically-nested edge detection (HED) [9] that performs image-to-image prediction by using fully convolutional neural networks and deeply supervised nets. In the cow motion classification process, we provide pose multiple linear models in which the coefficients of independent variables are utilized as features for classification.

The main contributions of the paper are

(i) The utilization of fully connected convolutional neural networks enables the proposed method to produce the edge map image as output;

(ii) The multiple linear model inspired by deeply-supervised nets gives us early classification results;

(iii) The favorable characteristics of the proposed holistically-nested edge detection technique are both accurate and computationally efficient;

(iv) The paper extends some application horizons of HED to wider and practical modern dairy farming.

The rest of the paper is organized as follows. Section 2 presents an overview of the proposed method. Some experimental results by using self-collected real-life data are shown in Section 3 including statistical analysis and discussions. Finally, we conclude the paper with concluding remarks in Section 4.

2. **Overview of the Proposed Method.** The general architecture of the proposed system is composed of two functional modules: edge-based cow region extraction module and motion classification module. The first module tackles two tasks: (i) the automatic detecting and segmentation of the cow's body region using HED method and (ii) extraction of bounding box for cow region based on You Only Look Once (YOLO) deep learning algorithm [19]. The second module is the training of deep convolutional neural network over the segmented cow's pattern images for the classification of the individual cow's movements such as lying, standing, transitions from lying to standing and vice-versa by using multiple linear models. According to the findings in [9], the holistically-nested edge detector (HED) can produce an edge map for an input image with an aid of deep layer supervision. The general architecture of the proposed system is shown in Figure 1.

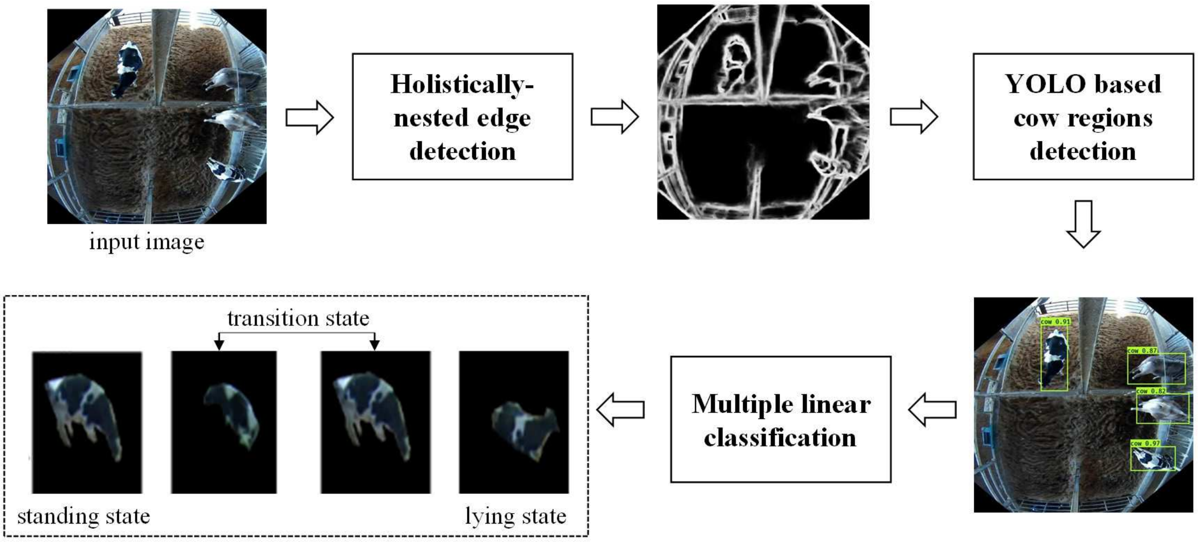

Figure 1. General architecture of the proposed system

2.1. **Holistically-nested edge-based cow region extraction module.** In this section, we describe the application of the holistically nested edge detection method introduced to extract edge images of dairy cow regions which can be utilized for analyzing cow behaviors. Briefly, the HED system can be considered as an end-to-end detection based on fully convolutional neural networks with additional deep supervision nets. More precisely, unlike the usual input training dataset, we have the forms of pairs (X, Y) where X denotes the raw input image and Y denotes the corresponding ground truth binary edge map for image X. The objective is to establish a network that learns features closely related to the ground truth. From background and input frames, edges are extracted by using HED. We then calculate the gray level absolute difference of two edge images. After that, we extract edge images by using thresholding where the threshold is taken as

$$Threshold = mean(GrayLevel) + std(GrayLevel) \times 0.1 \tag{1}$$

Finally, the edge regions of the cow body are extracted by using bounding boxes from YOLO algorithms. The flow diagram for the process is shown in Figure 2.

2.2. **Multiple linear model classification module.** After the extraction of cow regions, we shall analyze the behavior patterns of an individual cow. For the sake of simplicity, we limit ourself that at any instance time, a cow can be in the one of states standing, lying and transitions such as standing to lying and vice versa. However, the transition

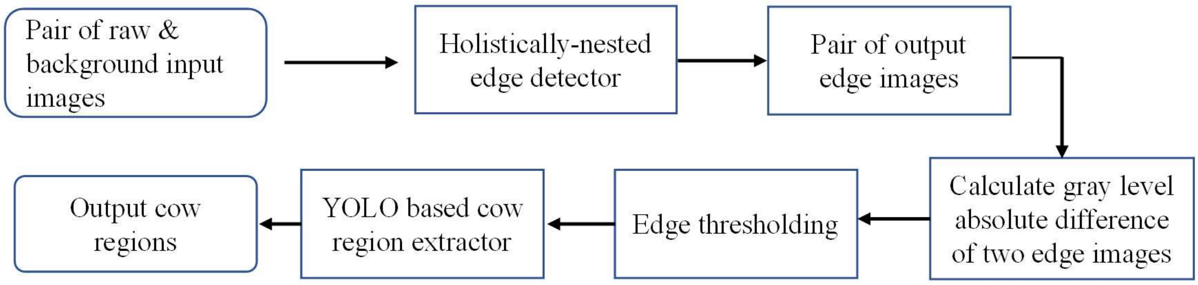

FIGURE 2. Process of holistically-nested edge-based cow region extraction module

states of lying to standing and standing to lying are significant for further cow behavior analysis for calving event prediction process. In this context, we shall investigate the patterns of how transition states occur during a certain period of times. In order to do so, we first extract the adjacent differences of centroid points from frame sequences while the cow is in motion.

We then extract the coordinate points (x, y) of centroid movement from a fixed number of frame sequences, and calculate maximum and minimum coordinate points by using the following definitions.

$$\begin{aligned}
&\min x = \textit{minimum} \ (x \text{ values in centroid movement sequence}) \\
&\min y = \textit{minimum} \ (y \text{ values in centroid movement sequence}) \\
&\max x = \textit{maximum} \ (x \text{ values in centroid movement sequence}) \\
&\max y = \textit{maximum} \ (y \text{ values in centroid movement sequence}) \\
&u = x - \text{distance} = \max(x) - \min(x) \\
&v = y - \text{distance} = \max(y) - \min(y)
\end{aligned}$$

We then define a multiple linear model of two independent variables u, v and dependent variable z representing the class labels. In our case there are only two transition classes say class 1 as standing to lying and class 2 as lying to standing.

Multiple linear model or multiple linear regression of u, v and z is defined as shown in Equation (2).

$$z = a + bu + cv \tag{2}$$

where a, b and c are to be determined by using sample statistical values of u, v and z in the training data sets. To make calculation easy, we use the following symbols and formulae.

$$\mu_x = \frac{\sum x}{N}, \text{ mean } x \text{ for } x = u, v, z, \ \sigma_x = \frac{\left(\sum x - \mu_x\right)^2}{N}, \text{ standard deviation } x \text{ for } x = u, v, z \tag{3}$$

Correlation coefficient of x and y

$$\rho(x, y) = \frac{\sum xy - N\mu_x\mu_y}{N\sigma_x\sigma_y} \tag{4}$$

We then have by using a little mathematical calculation:

$$b = \frac{\rho(u, z) - \rho(v, z)\rho(u, v)}{1 - (\rho(u, v))^2} * (\sigma_z/\sigma_u) \tag{5}$$

$$c = \frac{\rho(v, z) - \rho(u, z)\rho(u, v)}{1 - (\rho(u, v))^2} * (\sigma_z/\sigma_v) \tag{6}$$

$$a = \mu_z - b\mu_u - c\mu_v \tag{7}$$

We can then utilize Equation (2) to predict class label for the testing data. We shall perform some experiments for the class prediction by using real-life data collected from a large dairy farm in Japan.

3. Some Experimental Results.

3.1. Holistically-nested edge detection and YOLO experiment. In order to confirm our proposed holistically-nested edge detection method for cow region extraction and transition states classification, we collected video sequences continuously taken in some calving barns until the calving events occur. Out of the collected video sequences, the snapshots of frame sequences are prepared. The original dimensions of frames from overhead 360 cameras are (2560×1920) resolution. The size of frames is cropped and resized as the dimensions of (640×600) resolution for the purpose of reducing in calculation times. The dataset information is given in Table 1.

TABLE 1. Dataset information

Specification	Value
Video duration	5 minutes
Number of videos	30
Original image resolution	2560×1920 pixels
Resized image resolution	640×600 pixels
Extracted frame rate	1 fps
Length of analysis frames (1 minute)	60 frames

First, we employ the proposed holistically-nested edge detection method for edge image detection for our dataset. The sample results are shown in Figure 3.

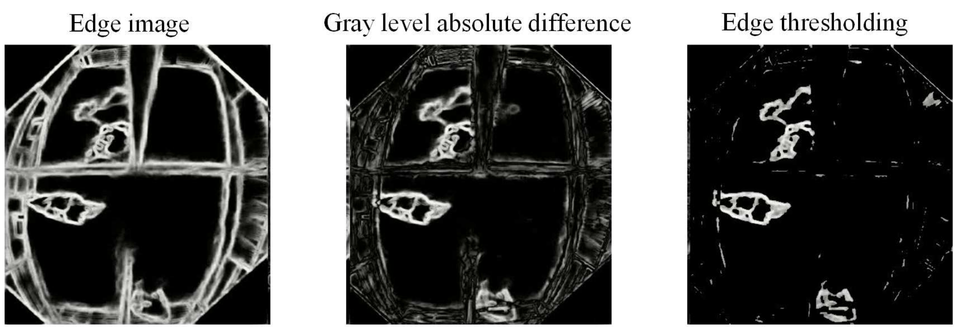

FIGURE 3. A sample result of HED based edge detection

Then, after employing YOLO detector [19], the results become as shown in Figure 4.

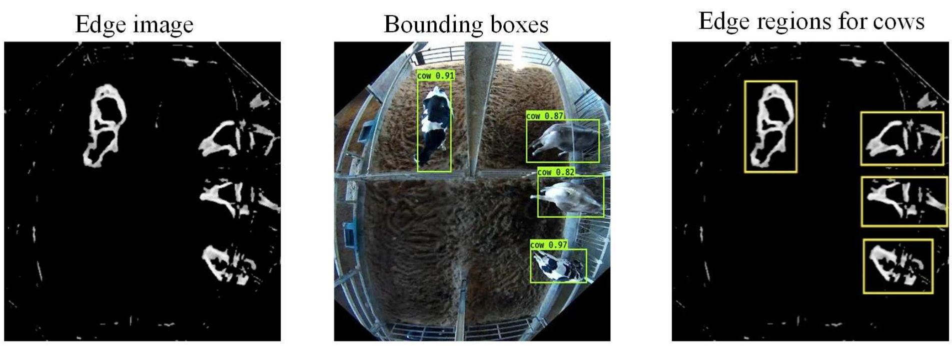

FIGURE 4. An illustration of edges extraction over cow regions

These results are tuned by using post processing so that we obtain as described in Figure 5.

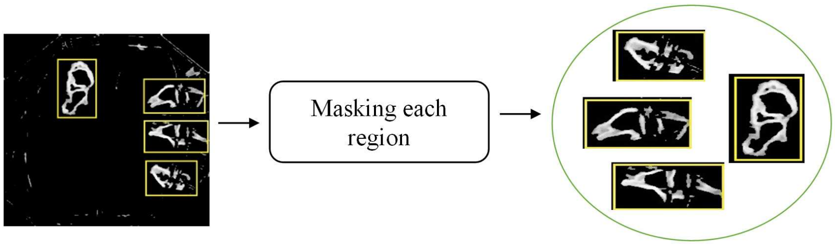

FIGURE 5. The foreground objects extraction by post processing

Thus, the final experimental sample result is seen as shown in Figure 6.

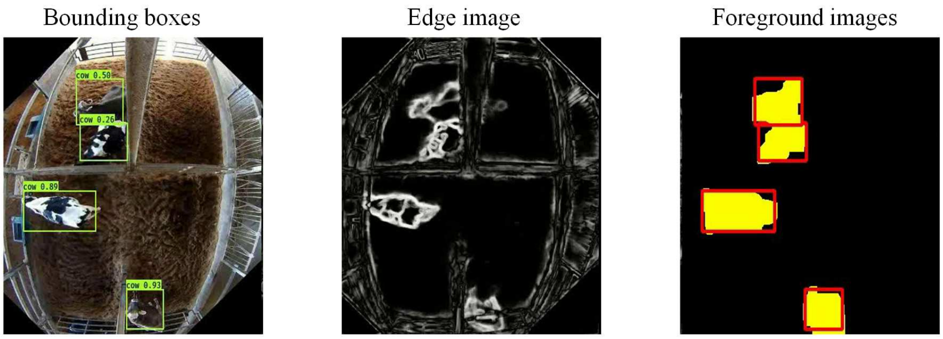

FIGURE 6. An illustration of final experimental results

3.2. **Multiple linear model classification experiment.** For this experiment, we analyze the centroid movements of detected cow regions extracted in the holistically-nested edge detection and YOLO experiment above. As a sample we implement on the dataset consisting of 51 snap shots (1 shot – 21 frames) in which there are 26 transitions of standing to lying labeled as class 1 and 25 transitions of lying to standing labeled as class 2.

We then calculate $u = x -$ distance, $v = y -$ distance of centroid movements and class label $z = 1$ or 2. They are described in Table 2.

From 51 shots in the dataset, we use 30 for training and 21 for testing. For the 21 dataset, the accuracy calculation is done. We have tested 21 video sequences in our own datasets, and it is found that 19 are correctly classified. Therefore, we just perform a simple accuracy calculation such that (corrected number divided by total number) which gives us $(19/20) = 0.904762 = 90.48\%$.

We achieve an accuracy of 90.48%. We also made comparison with other methods such as rule based and support vector machine methods.

In the rule-based method, first define

$$\beta_1 = \operatorname{cov}(x, t)/\sigma_x^2 \tag{8}$$

$$\beta_2 = \operatorname{cov}(y, t)/\sigma_y^2 \tag{9}$$

where x and y are coordinates of extreme points and t stands for a frame number in the dataset. σ_x and σ_y are the standard deviations of x sequence and y sequence respectively.

We then define the beta rule as follows:

(i) if $\beta_1 > 0$ and $\beta_2 > 0$ then trasition state of "Lying to Standing"
(ii) if $\beta_1 > 0$ and $\beta_2 < 0$ then no posture changes
(iii) if $\beta_1 < 0$ and $\beta_2 < 0$ then trasition state of "Standing to Lying"
(iv) if $\beta_1 < 0$ and $\beta_2 > 0$ then no posture changes

Table 2. Dataset information

No.	u	v	z	No.	u	v	z	No.	u	v	z
1	−2.48	−0.72	1	18	0.06	−1.38	1	35	1.55	0.03	2
2	−2.45	−0.36	1	19	−3.37	0.39	1	36	0.62	0.03	2
3	−2.17	−0.48	1	20	−0.58	−0.39	1	37	3.07	−0.57	2
4	−1.82	0.12	1	21	−2.86	−0.83	1	38	2.5	2.42	2
5	−1.73	−0.91	1	22	−2.8	−0.45	1	39	−1.07	0.57	2
6	0.39	−0.84	1	23	−0.55	−0.74	1	40	1.27	1.96	2
7	−3.19	−0.47	1	24	−2.27	−0.84	1	41	4.08	0.82	2
8	−1.03	−0.21	1	25	−2.42	−1.43	1	42	1.19	1.4	2
9	−0.38	−1.17	1	26	−1.7	−0.96	1	43	1.7	0.2	2
10	−1.34	−0.04	1	27	1.42	0.68	2	44	0.5	0.55	2
11	0.09	−0.32	1	28	1.46	0.76	2	45	1.64	0.04	2
12	−1.66	−0.88	1	29	2.56	1.58	2	46	1.84	−0.57	2
13	−2.87	0.17	1	30	1.96	1.14	2	47	2.69	0.78	2
14	−1.32	−0.6	1	31	1.3	1.89	2	48	1.84	2.34	2
15	1.53	−0.16	1	32	2.75	0.71	2	49	3.91	−0.44	2
16	−1.73	−0.91	1	33	2.57	−0.11	2	50	1.96	0.38	2
17	−1.26	0.34	1	34	3.81	0.42	2	51	0.58	1.33	2

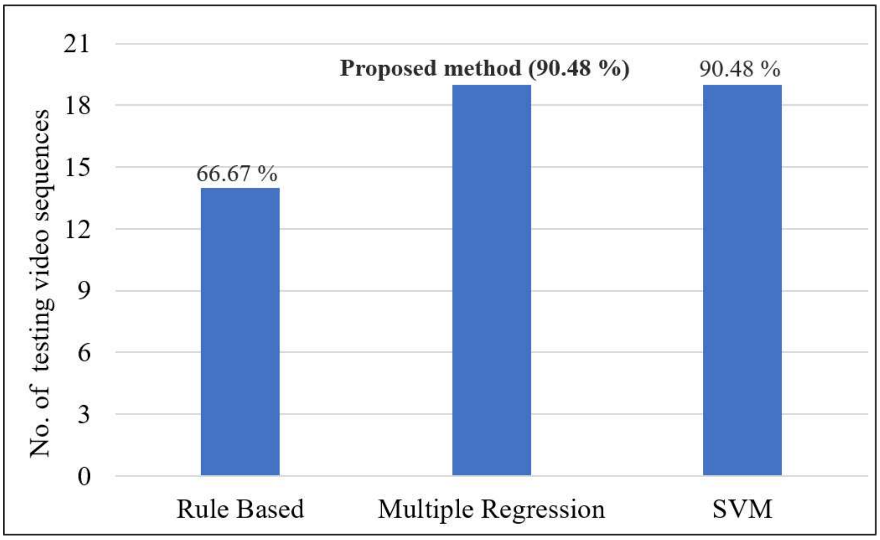

Figure 7. Accuracy analysis for the proposed method

By applying this rule for dataset of 21 video sequences only 14 are correctly classified. This will give an accuracy of $0.666666667 = 67\%$. In similar fashion, we test the same dataset by using SVM classification, and the accuracy is as shown in Figure 7.

4. **Conclusions.** In this paper we had proposed holistically-nested edge-based detector for dairy cow region extraction in conjunction with YOLO deep learning algorithms. Using the extracted cow regions, we investigated transition behaviors of cow motions that can be utilized to predict time to calving events. Among many others we noticed that the states of transitions are significantly relied on the amounts of centroid movements. However, it may need more experiments to be done. We hope these problems would be performed in our future work.

Acknowledgments. This work was supported in part by Honkawa Ranch Research Grant 2019-A-01. We thank K. Honkawa, D.V.M. and Professor Y. Horii for giving every convenience of the study in the ranch and their valuable advice.

REFERENCES

[1] K. Zhao, X. Jin, J. Ji, J. Wang, H. Ma and X. Zhu, Individual identification of Holstein dairy cows based on detecting and matching feature points in body images, *Biosystems Engineering*, vol.181, pp.128-139, 2019.

[2] W. Shao, R. Kawakami, R. Yoshihashi, S. You, H. Kawase and T. Naemura, Cattle detection and counting in UAV images based on convolutional neural networks, *International Journal of Remote Sensing*, vol.41, no.1, pp.31-52, 2020.

[3] T. T. Zin, M. Z. Pwint, P. T. Seint, S. Thant, S. Misawa, K. Sumi and K. Yoshida, Automatic cow location tracking system using ear tag visual analysis, *Sensors*, vol.20, no.12, DOI: 10.3390/s20123564, 2020.

[4] L. J. Fleishman and J. A. Endler, Some comments on visual perception and the use of video playback in animal behavior studies, *Acta Ethologica*, vol.3, no.1, pp.15-27, 2000.

[5] M. Zabłocki, K. Gościewska, D. Frejlichowski and R. Hofman, Intelligent video surveillance systems for public spaces – A survey, *Journal of Theoretical and Applied Computer Science*, vol.8, no.4, pp.13-27, 2014.

[6] M. Vrigkas, C. Nikou and I. A. Kakadiaris, A review of human activity recognition methods, *Frontiers in Robotics and AI*, vol.2, DOI: 10.3389/frobt.2015.00028, 2015.

[7] T. T. Zin, C. N. Phyo, P. Tin, H. Hama and I. Kobayashi, Image technology based cow identification system using deep learning, *Proc. of the International MultiConference of Engineers and Computer Scientists*, vol.1, pp.236-247, 2018.

[8] H. Liu, A. R. Reibman and J. P. Boerman, A cow structural model for video analytics of cow health, *arXiv Preprint*, arXiv: 2003.05903, 2020.

[9] S. Xie and Z. Tu, Holistically-nested edge detection, *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp.1395-1403, DOI: 10.1109/ICCV.2015.164, 2015.

[10] K. K. Maninis, S. Caelles, Y. Chen, J. Pont-Tuset, L. Leal-Taixé, D. Cremers and L. Van Gool, Video object segmentation without temporal information, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.41, no.6, pp.1515-1530, 2018.

[11] A. Poursaberi, C. Bahr, A. Pluk, A. Van Nuffel and D. Berckmans, Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques, *Computers and Electronics in Agriculture*, vol.74, no.1, pp.110-119, 2010.

[12] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada and Y. Yagi, RGB-D video-based individual identification of dairy cows using gait and texture analyses, *Computers and Electronics in Agriculture*, vol.165, DOI: 10.1016/j.compag.2019.104944, 2019.

[13] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis and M. Bethge, DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning, *Nature Neuroscience*, vol.21, no.9, pp.1281-1289, 2018.

[14] K. Zhao, J. M. Bewley, D. He and X. Jin, Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique, *Computers and Electronics in Agriculture*, vol.148, pp.226-236, 2018.

[15] A. Ter-Sarkisov, R. Ross and J. Kelleher, Bootstrapping labelled dataset construction for cow tracking and behavior analysis, *The 14th Conference on Computer and Robot Vision (CRV)*, pp.277-284, 2017.

[16] K. He, G. Gkioxari, P. Dollár and R. Girshick, Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.2961-2969, 2017.

[17] T. T. Zin, S. Z. M. Maung, P. Tin and Y. Horii, Feature detection and analysis of cow motion classification for predicting calving time, *International Journal of Biomedical Soft Computing and Human Sciences (IJBSCHS)*, vol.26, no.1, pp.11-20, 2021.

[18] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.40, no.4, pp.834-848, 2017.

[19] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *arXiv Preprint*, arXiv: 1804.02767, 2018.