International Journal of Innovative Computing, Information and Control
Volume **20**, Number **5**, October **2024**

ICIC International ©2024 ISSN 1349-4198
pp. 1439–1447

AUTOMATED CATTLE DETECTION USING MASK R-CNN AND IOU-BASED TRACKING WITH A SINGLE SIDE-VIEW CAMERA

Bo Bo Myint[1], Tsubasa Onizuka[1], Pyke Tin[1], Masaru Aikawa[2]
Ikuo Kobayashi[3] and Thi Thi Zin[1,*]

[1]Graduate School of Engineering
[2]Organization for Learning and Student Development
[3]Sumiyoshi Livestock Science Station, Field Science Center, Faculty of Agriculture
University of Miyazaki
1-1, Gakuen kibanadai-Nishi, Miyazaki 889-2192, Japan
{ z322702; hl18012 }@student.miyazaki-u.ac.jp; pyketin11@gmail.com
{ aikawa; ikuokob }@cc.miyazaki-u.ac.jp
*Corresponding author: thithi@cc.miyazaki-u.ac.jp

Received December 2023; revised March 2024

Abstract. *In precision livestock farming, the early detection of lameness in cattle is an extremely important aspect of effective breeding management. Timely identification of lameness not only facilitates prompt and cost-efficient treatment but also plays a crucial role in avoiding possible future diseases. This study emphasizes the significance of intelligent visual perception systems for lameness detection in dairy cattle, particularly in the lane between from Milking Parlor to Cattle Barn. To address the cattle lameness issue, we employ an advanced deep learning, and image processing technique, i.e., Mask R-CNN from Detectron2 to detect and identify cattle regions for feature extraction of lameness detection. On the other hand, cattle tracking using IoU is also an important part of data accumulation for lameness classification. The results of this study contribute to ongoing efforts in precision animal husbandry and demonstrate the potential of intelligent visual recognition systems for early lameness detection.*
Keywords: Cattle detection, Cattle tracking, Detectron2, Mask R-CNN

1. **Introduction.** In precision livestock production, lameness detection in cattle is especially important for reproductive management [1]. Among dairy farming, lameness has a significant impact on the cattle's reproductive performance and milk production. For herd productivity and animal welfare, lameness has become a frequent and serious problem. This is because pain, hoof and joint discomfort, and the weight and mobility of the lame cow prevent normal behavior [2]. To provide effective treatment and prevention of future illness, early lameness detection becomes the important factor. For the detection of the lameness of the cattle, automatic lameness detection systems such as sensor-based systems and vision-based systems are investigated [3]. Lameness detection is one of the challenges of the dairy industry because lameness is the world's third rank and most costly health problem of dairy cattle [4]. By applying the traditional method, manually scoring the lameness by the human through visual observation can be time consuming, labor intensive and costly. Nowadays, the lameness detection by using the computer vision is increased [5]. To determine the level of lameness in cattle, the gait and posture of the animals were considered. Current automated lameness detection methods, ranging from

DOI: 10.24507/ijicic.20.05.1439

leg-load distribution systems to image processing-based approaches, hold significant potential for precise detection by monitoring various cattle characteristics, offering benefits such as enhanced welfare, reduced labor costs, and increased productivity in the dairy industry [6]. The meta-analysis of British dairy cattle indicates a 29.5% pooled prevalence of lameness, with an all-cause incidence rate of 30.9 cases per 100 cattle, revealing diversity in data collection methods and underscoring the necessity for standardization in addressing this persistent issue [7]. A camera-based lameness monitoring algorithm for each individual cattle in a large dairy herd, demonstrating high specificity, sensitivity, accuracy, and precision by utilizing historical data and individualized thresholds [8]. An automatic lameness detection tool utilizing low-frequency acceleration signals from leg-worn sensors on cattle, employing a triaxial accelerometer and a time-frequency-based Long Short-Term Memory (LSTM) neural network to accurately identify gait phases, estimate frequencies, enhance movement coherency, and demonstrate versatility across various applications [9]. The considerable impact of cattle lameness on dairy cattle and the industry, presenting a lameness detection method using pre-trained neural networks on annotated video clips, with an emphasis on cattle structure through binary segmentation masks for potential early detection on farms [10]. A comprehensive overview of the current state and prospects future of computer vision techniques for detecting lameness in dairy cattle, addressing challenges, and emphasizing the need for improved accuracy and applicability despite the method's noncontact advantages and affordability [11]. In [12], an RGB camera is utilized for analyzing the walking behavior of the cattle. From [13], the Mask R-CNN is efficient in the usage of cattle detection of cattle mounting analysis.

Various levels of cattle lameness, ranging from 1 to 4, were observed in our dataset collected from the farm as shown in Figure 1. Before lameness classification, our system employs advanced computer vision technology to detect and confirm the exact location of cattle. Recognizing the fundamental importance of accurately locating cattle, we integrate two state-of-the-art algorithms, Mask R-CNN [14] from Detectron2, to effectively identify cattle regions within the visual field. To ensure robust tracking across frames and maintain continuity of cattle identification, Intersection over Union (IoU) calculations are applied; the IoU metric serves as a pivotal tool, quantifying the spatial overlap between detected regions and allowing the monitoring process. With the integration of the Detectron2 and the incorporation of IoU calculations, our system establishes a solid foundation for reliable cattle detection and cattle tracking before lameness classification in the context of precision livestock farming. The accurate cattle detection and tracking results for those that pass through the lane are depicted in Figure 2.

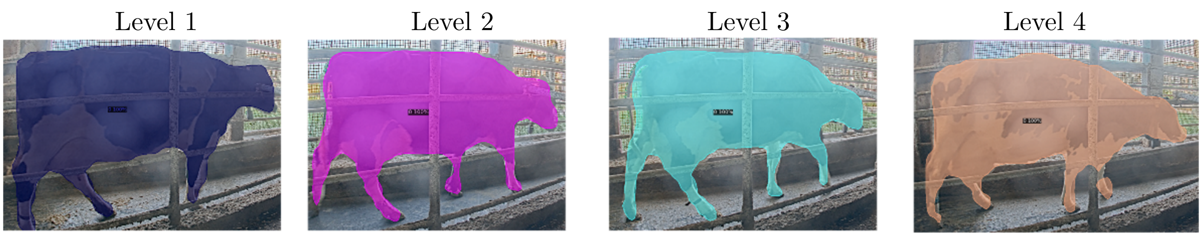


FIGURE 1. Sample lameness levels from farm

2. **Proposed Methodology.** The proposed system consists of three components: data preprocessing, detection, and tracking, as shown in Figure 3. The first step is data preprocessing, where frames are extracted from the video source at a rate of one frame per second (1 fps). The Detectron2 algorithm is then employed in the detection process, resulting in

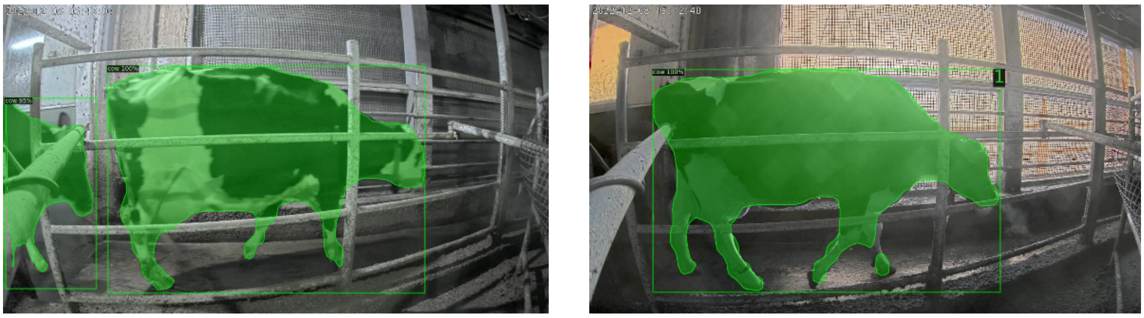


FIGURE 2. Cattle detection and cattle tracking

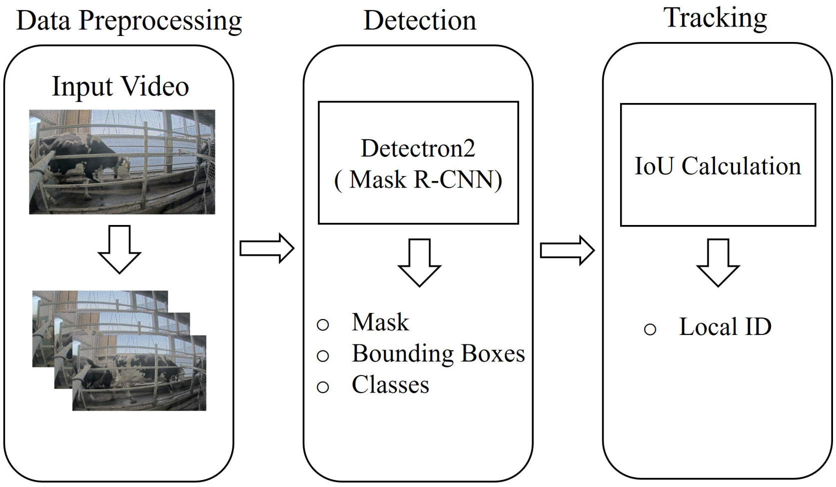


FIGURE 3. Overall proposed system of cattle lameness system

three key features (masked regions, bounding boxes, and cow classes). In particular, the focus of the detection phase is to exploit the mask region features. The tracking phase utilizes Intersection over Union (IoU) computations for local IDs, contributing to the system's ability to efficiently track identified cow instances across frames. This modular and systematic approach ensures that the appropriate information is extracted from the video source, facilitates robust detection and tracking of cows, and lays the foundation for subsequent analyses such as lameness classification.

2.1. **Data preprocessing.** In the context of this research investigation, the system is operated the datasets originating from the Hokuren Kunneppu Demonstration Farm, situated in the Hokkaido prefecture of Japan. Within the cattle farm infrastructure, two dedicated lanes serve as conduits between the Milking parlor and the cattle barn. Since Lane A and Lane B have exactly similar structures, only Lane A was used in the proposed system. At the data acquisition, a sophisticated 4K camera, specifically the AXIS P 1448-LE, has been strategically deployed at the initiation point of Lane A, as delineated in Figure 4. This high-resolution camera, boasting dimensions of 3,840 $\times$ 2,160 pixels, is selected with precision to capture imagery of exceptional detail and quality. Notably, the original frame rate of 25 frames per second (fps) necessitates pragmatic adjustments, with the system calibrated to a reduced 1 fps for optimal data annotation processes. In data annotation for the training of the cattle detection model, the dataset utilized is comprehensively outlined in Table 1. A total of 3,000 images from 56 cattle, captured on July 04, 2022, during both morning and evening sessions, have been meticulously selected for the purpose of training the cattle detection pipeline.

FIGURE 4. Camera setting from Lane A of cattle farm

TABLE 1. Dataset used for training of cattle detection

Date	Duration	#Images
July 04, 2022	Morning, Evening	3,000

2.2. **Cattle detection.** In the domain of cattle detection, the annotated images undergo optimization for training purposes through the utilization of Mask R-CNN from Detectron2, a framework developed by Facebook AI Research (FAIR) to expedite the implementation and evaluation of innovative computer vision research. To streamline the training process and optimize GPU utilization, the original frame size is reduced to 1,280 $\times$ 720. The core methodology centers around the application of Mask R-CNN instance segmentation, a technique paramount in extracting precise mask regions corresponding to cattle within real-world, real-time scenarios. The trained weight file derived from this process is then employed on input videos, generating binary mask regions of cattle. These binary masks are pivotal in subsequent feature extraction for cattle lameness calculation and determining bounding boxes crucial for the tracking and identification of individual cattle. The original image is preserved to accommodate future processes pertaining to feature extraction and detection. The comprehensive detection process is visually depicted in Figure 5, offering a representation of steps involved in the overall detection methodology.

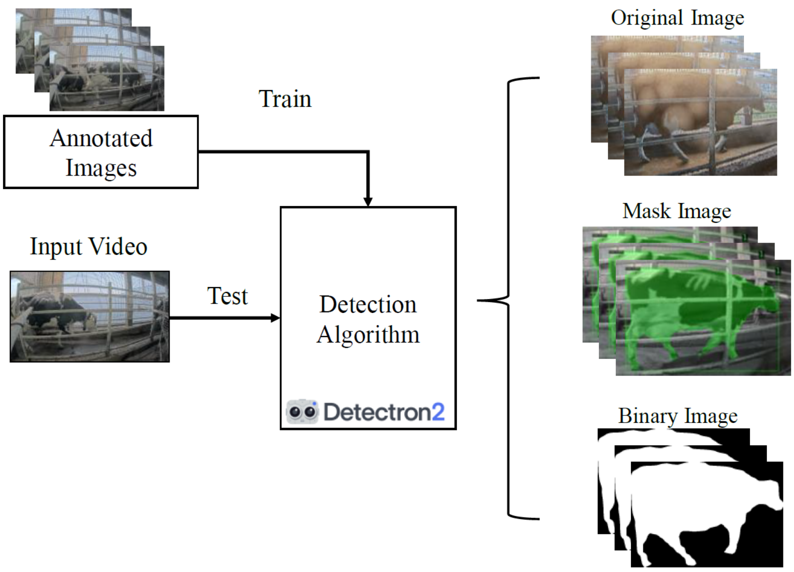


FIGURE 5. Cattle detection process

2.3. **Cattle tracking.** In the cattle tracking phase, the determination of bounding boxes for each cattle is executed through the application of Intersection over Union (IoU), a metric that quantifies the degree of overlap between two bounding boxes encompassing cattle within consecutive frames. The IoU, reflective of the extent of intersection relative to the union of the bounding boxes, serves as a key criterion for assessing the spatial congruence of detected cattle. In this research, the IoU computation spans both preceding and current frames, enabling the establishment of a unique identification number for each cow based on the detection outcomes. Relative to alternative tracking algorithms, the employment of IoU calculation between previous and current frames emerges as a more efficient and effective approach in terms of runtime and resource utilization. A cohort exceeding 50 to 60 cattle is meticulously delineated and accurately separated during the tracking phase, showcasing the efficacy of the tracking algorithm. The results of cattle detection, each associated with a unique local identification number, are systematically organized into folders, as illustrated in Figure 6. Following the completion of the tracking phase, the system's outcomes become instrumental for future computations in the context of lameness classification. The equation used for IoU calculations is shown as

$$IoU = Area\ of\ Overlap/Area\ of\ Union \tag{1}$$

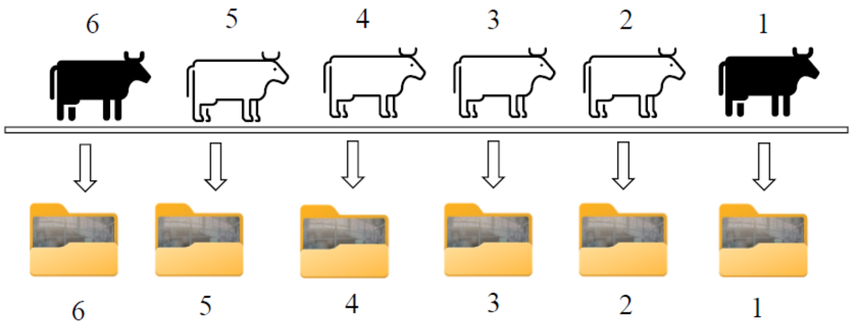


FIGURE 6. The result of cattle tracking process

3. **Experimental Results.** In this section, the cattle detection results obtained through Mask R-CNN from Detectron2 and the cattle tracking results using IoU on the testing dataset of December 2022 are presented as follows.

3.1. **Cattle detection results.** For efficient detection results of cattle mask regions, the widely used Mask R-CNN algorithm from the Detectron2 model was trained with iterations at 270, 300, and 330. The model trained with 300 iterations demonstrated superior performance in the cow detection system, outperforming the others based on the results from testing videos from December Datasets. Moreover, precision and recall values of the trained model were computed, and the equations used for calculating precision, recall, and accuracy are presented as follows:

$$Recall = TP/(TP + FN) \tag{2}$$

$$Precision = TP/(TP + FP) \tag{3}$$

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{4}$$

In the above equation, TP (True Positive) signifies correctly detected cattle instances, while FN (False Negative) denotes cases where exiting cattle instances were not detected. FP (False Positive) occurs when the background is incorrectly identified as the cattle region, with one of the moving object's human regions being considered as part of the background. TN (True Negative) represents situations where there are no cattle, and indeed, there are no cattle. Consequently, many object detection methodologies do not account for True Negative, and this research will not calculate the True Negative. The number of instances varies based on the quantity of cattle passing through. The instances range from crowded situations with traffic, where cattle may stand in front of the camera, to cases where cattle pass through swiftly. In the December dataset, the recall results for all testing days are consistently 100%, indicating no missed cattle regions in the detection. However, the accuracy varies based on the rate of incorrect detection in the human region. The average accuracy across all six testing datasets is 98.21%.

Table 2. Dataset used for training of cattle detection

Date	**#Instances**	**#Correct Instances (TP)**	**#Incorrect Instances (FP)**	**#Miss Instances (FN)**	**Recall (%)**	**Precision (%)**	**Accuracy (%)**
Dec 6, 2022 (M)	4,810	4,532	278	0	100.00	94.22	94.22
Dec 6, 2022 (E)	4,233	4,219	14	0	100.00	99.67	99.67
Dec 7, 2022 (M)	4,856	4,748	108	0	100.00	97.78	97.78
Dec 7, 2022 (E)	7,057	7,019	38	0	100.00	99.46	99.46
Dec 8, 2022 (M)	11,248	11,139	109	0	100.00	99.03	99.03
Dec 8, 2022 (E)	6,627	6,570	57	0	100.00	99.13	99.13

Here M means morning, and E for evening.

3.2. **Cattle tracking results.** In this section, cattle tracking results involve IoU calculation between cattle instances in previous and current frames, with each cattle grouped into separate folders identified locally from 1 to N, where N represents the number of cattle passing through the session. Table 3 displays the results for December 6, 7, and 8, during both morning and evening periods. While the tracking accuracy for all datasets on December 6 and 7, in both morning and evening, reaches 100%, the accuracy for the December 8 dataset is approximately 96%, attributed to an overcrowded situation with the cattle. The collective average accuracy for the December datasets surpasses 98.94%. The illustrated sample cattle tracking result is presented in Figure 7, where the local identification number is indicated on the rightmost part of the bounding box surrounding the detected cattle region.

Table 3. Dataset used for training of cattle detection

Date	**#Cattle**	**#Correct Cattle**	**#Wrong Cattle**	**Accuracy (%)**
Dec 6, 2022 (M)	58	58	0	100.00
Dec 6, 2022 (E)	58	58	0	100.00
Dec 7, 2022 (M)	51	51	0	100.00
Dec 7, 2022 (E)	58	58	0	100.00
Dec 8, 2022 (M)	66	64	2	96.97
Dec 8, 2022 (E)	60	58	2	96.67

Here M means morning, and E for evening.

4. **Conclusion.** In conclusion, this study introduces an efficient and accurate cattle detection system leveraging Mask R-CNN from Detectron2 for precise cow region identification. Notably, the tracking process in the proposed system stands out for its cost-efficiency and time savings, relying solely on Intersection over Union (IoU) calculations. The results

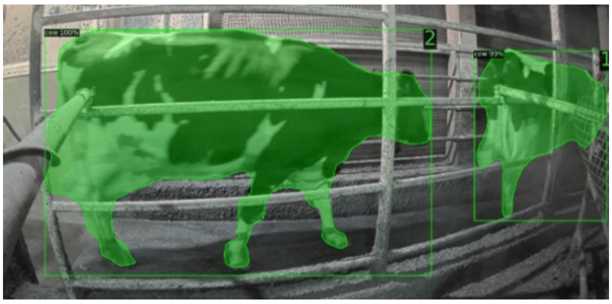


Figure 7. Cattle tracking result

obtained from the tracking and detection processes yield valuable information crucial for the subsequent calculation of lameness classification. The proposed system exhibits commendable performance, achieving an average accuracy of 98.21% for cattle mask region detection in the December testing datasets. Additionally, the average accuracy for cattle tracking in the December datasets surpasses 98.94%. The future trajectory of this research involves expanding the scope by applying various deep learning techniques to larger datasets, with an emphasis on accuracy comparison. Furthermore, diverse deep learning algorithms will be explored to address the challenge of overlapping cows passing through lanes individually, aiming for enhanced accuracy in cow tracking. This study contributes to the ongoing advancements in precision livestock farming, offering a robust framework for cattle detection and tracking with promising avenues for further refinement and application. In the future, the reliability of these cattle detection and tracking results will be leveraged for advanced regional feature extraction, enabling the calculation of cattle lameness classification and identification. This approach aims to facilitate the monitoring of cattle health care systems in real-world applications.

Acknowledgment. We would like to thank the people at the Kunneppu Demonstration Farm for the great accommodation and valuable advice for our research at the farm. This publication was subsidized by JKA through its promotion funds from KEIRIN RACE.

REFERENCES

[1] S. M. Noe et al., A deep learning-based solution to cattle region extraction for lameness detection, *2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, 2022.

[2] Y. Karoui, A. B. Jacques, A. Diallo, E. Shepley and E. Vasseur, A deep learning framework for improving lameness identification in dairy cattle, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol.35, pp.15811-15812, DOI: 10.1609/aaai.v35i18.17902, 2021.

[3] 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.

[4] T. T. Zin, Y. Htet, S. C. Tun and P. Tin, Artificial intelligence topping on spectral analysis for lameness detection in dairy cattle, *Proceedings of the Annual Conference of Biomedical Fuzzy Systems Association*, vol.35, C-3, DOI: 10.24466/pacbfsa.35.0_C-3, 2022.

[5] S. Han, J. Zhang, G. Cheng, Y. Peng, J. Zhang and J. Wu, Current state and challenges of automatic lameness detection in dairy cattle, *Smart Agriculture*, vol.2, no.3, pp.21-36, 2020.

[6] S. Barney, S. Dlay, A. Crowe, I. Kyriazakis and M. Leach, Deep learning pose estimation for multi-cattle lameness detection, *Scientific Reports*, vol.13, no.1, DOI: 10.1038/s41598-023-31297-1, 2023.

[7] J. Afonso, M. Bruce and P. Keating, Profiling detection and classification of lameness methods in British dairy cattle research: A systematic review and meta-analysis, *Frontiers in Veterinary Science*, vol.7, 542, DOI: 10.3389/fvets.2020.00542, 2020.

[8] D. Piette, T. Norton, V. Exadaktylos and D. Berckmans, Individualised automated lameness detection in dairy cows and the impact of historical window length on algorithm performance, *Animal*, vol.14, no.2, pp.409-417, 2020.

[9] D. Jarchi, J. Kaler and S. Sanei, Lameness detection in cows using hierarchical deep learning and synchrosqueezed wavelet transform, *IEEE Sensors Journal*, vol.21, no.7, pp.9349-9358, 2021.

[10] E. Arazo, R. Aly and K. McGuinness, Segmentation enhanced lameness detection in dairy cows from RGB and depth video, *arXiv Preprint*, arXiv: 2206.04449, 2022.

[11] X. Kang, X. D. Zhang and G. Liu, A review: Development of computer vision-based lameness detection for dairy cows and discussion of the practical applications, *Sensors*, vol.21, no.3, 753, DOI: 10.3390/s21030753, 2021.

[12] T. Onizuka, T. T. Zin and I. Kobayashi, A non-invasive method for lameness detection in dairy cows using RGB cameras, *ICIC Express Letters, Part B: Applications*, vol.14, no.10, pp.1107-1114, DOI: 10.24507/icicelb.14.10.1107, 2023.

[13] 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, DOI: 10.24507/ijicic.18.01.211, 2022.

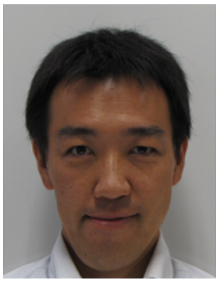
[14] K. Dollár and R. Girshick, Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, 2017.

Author Biography

Bo Bo Myint received the B.Sc. degree in Computer Science from the Department Information Science, University of Computer Studies, Yangon, Myanmar, in 2018. He is currently studying in the Master's degree Program of the Graduate School of Engineering, University of Miyazaki, Miyazaki, Japan. His research interests include computer vision, machine learning and deep learning.

Tsubasa Onizuka received the B.E. degree in Electrical Systems Engineering from the Department of Engineering, University of Miyazaki, Miyazaki, Japan, in 2022. He is currently studying in the Master's degree Program of the Graduate School of Engineering, University of Miyazaki, Miyazaki, Japan. His research interests include computer vision and image processing techniques.

Pyke Tin received the B.Sc. degree (Hons.) in Mathematics from the University of Mandalay, Myanmar, in 1965, the M.Sc. degree in Computational Mathematics from the University of Rangoon, Myanmar, in 1970, and the Ph.D. degree in Stochastic Processes and Their Applications from Monash University, Australia, in 1976. He was a Rector of the University of Computer Studies, Yangon, and a Professor of computational mathematics. He is currently a Visiting Professor with the International Relations Center, and also the Distinguished Professor of Faculty of Engineering, University of Miyazaki, Miyazaki, Japan. His research interests include image search engines, queuing systems, computer vision, stochastic processes, and their applications to image processing.



Masaru Aikawa received the B.Sc. degree in 2007 in Hokkaido Information University. He is an Assistant Professor of Organization for Learning and Student Development, University of Miyazaki, Miyazaki, Japan. His research interests include networking, deep learning and educational support system.

Ikuo Kobayashi received the D.V.M. degree in Veterinary Medicine from University of Miyazaki, Miyazaki, Japan in 1996. He received his Ph.D. degree 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.

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 the University of Computer Studies, Yangon, Myanmar. She received her Master's 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 at the 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.