

Article

Optimizing Cattle Behavior Analysis in Precision Livestock Farming: Integrating YOLOv7-E6E with AutoAugment and GridMask to Enhance Detection Accuracy

Hyeon-seok Sim ¹, Tae-kyeong Kim ¹, Chang-woo Lee ², Chang-sik Choi ², Jin Soo Kim ³ and Hyun-chong Cho ^{1,4,*} 

¹ Department Graduate Program for BIT Medical Convergence, Kangwon National University, Chuncheon 24341, Republic of Korea

² Gangwon State Livestock Research Institute, Hoengseong-gun 25266, Republic of Korea

³ College of Animal Life Sciences, Kangwon National University, Chuncheon 24341, Republic of Korea

⁴ Department of Electronics Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea

* Correspondence: hyuncho@kangwon.ac.kr

Abstract: Recently, the growing demand for meat has increased interest in precision livestock farming (PLF), wherein monitoring livestock behavior is crucial for assessing animal health. We introduce a novel cattle behavior detection model that leverages data from 2D RGB cameras. It primarily employs you only look once (YOLO)v7-E6E, which is a real-time object detection framework renowned for its efficiency across various applications. Notably, the proposed model enhances network performance without incurring additional inference costs. We primarily focused on performance enhancement and evaluation of the model by integrating AutoAugment and GridMask to augment the original dataset. AutoAugment, a reinforcement learning algorithm, was employed to determine the most effective data augmentation policy. Concurrently, we applied GridMask, a novel data augmentation technique that systematically eliminates square regions in a grid pattern to improve model robustness. Our results revealed that when trained on the original dataset, the model achieved a *mean average precision (mAP)* of 88.2%, which increased by 2.9% after applying AutoAugment. The performance was further improved by combining AutoAugment and GridMask, resulting in a notable 4.8% increase in the *mAP*, thereby achieving a final *mAP* of 93.0%. This demonstrates the efficacy of these augmentation strategies in improving cattle behavior detection for PLF.

Keywords: AutoAugment; cattle behavior; deep learning; GridMask; precision livestock farming; object detection; YOLOv7-E6E



Citation: Sim, H.-s.; Kim, T.-k.; Lee, C.-w.; Choi, C.-s.; Kim, J.S.; Cho, H.-c. Optimizing Cattle Behavior Analysis in Precision Livestock Farming: Integrating YOLOv7-E6E with AutoAugment and GridMask to Enhance Detection Accuracy. *Appl. Sci.* **2024**, *14*, 3667. <https://doi.org/10.3390/app14093667>

Academic Editors: Antonio Fernández-Caballero and Gianluigi Ferrari

Received: 5 February 2024

Revised: 8 April 2024

Accepted: 22 April 2024

Published: 25 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid increase in the global population has resulted in an increased demand for beef [1]. Figure 1, published by the Organization for Economic Co-operation and Development (OECD), shows beef consumption per capita by OECD countries during 2014–2021 [2], wherein it is evident that beef consumption has been increasing annually owing to the rising demand.

Therefore, precision livestock farming (PLF) is actively being developed to ensure effective and efficient livestock production [3]. PLF involves the development of monitoring systems for various livestock characteristics, including health and welfare [4]. Analyzing livestock behavior in PLF is crucial because it enables the assessment of the current conditions of animals [5]. Detecting the behavior of livestock not only aids in the real-time monitoring of animal health but also plays a crucial role in farm management. The ability to detect changes in behavior can act as an early indicator of potential health issues, enabling timely intervention and treatment. Additionally, it can assist in quickly identifying and addressing any stressors or discomforts experienced by the livestock. In particular, observing behaviors such as drinking, eating, and lying down can help determine the health status of

livestock or diagnose diseases early [6,7]. However, monitoring livestock behavior requires substantial human resources and labor. Furthermore, the direct observation of livestock is unsustainable, as the observer's attention inevitably decreases because of fatigue. Therefore, recent studies have focused on methods that enable observing livestock behavior without the need for continuous human monitoring.

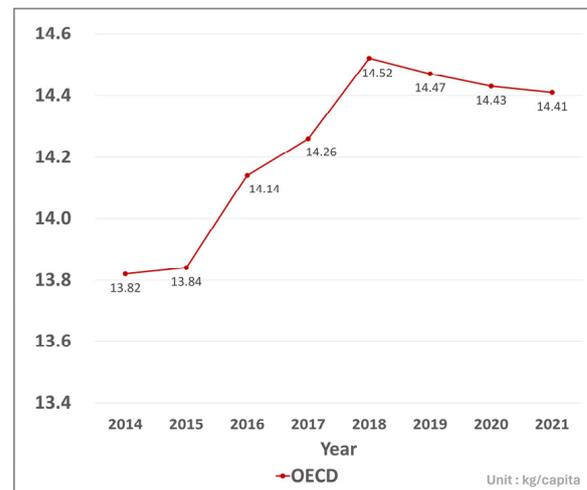


Figure 1. Beef consumption in OECD countries.

Pavlovic et al. [8] collected cow behavioral data through a neck-mounted three-axis accelerometer sensor and a jaw-mounted pressure sensor. Based on these data, they classified cow behaviors into three categories: ruminating, feeding, and others. Their classification model employed a convolutional neural network architecture and achieved an F1 score of 0.82, which indicates the good overall accuracy of the model. Similarly, Williams et al. [9] classified “defecation” and “urination” behaviors using an accelerometer attached to the cows’ tails. They developed a classification model using the random forest algorithm, which is an ensemble algorithm that makes predictions by learning from numerous decision trees, and achieved *sensitivity (recall)* and *precision* scores of >86.7%. Methods that analyze cattle behavior based on sensor data are effective and exhibit high performance for livestock behavior classification. However, such sensor-based data collection methods necessitate the direct attachment of sensors to livestock, which can cause stress in the animals. Additionally, these sensors are susceptible to damage and contamination; therefore, farmers must periodically check them, which increases their workload [10,11].

In contrast to research based on sensor-based methods, some studies have determined livestock behavior using data obtained from 3D cameras. Chen et al. [12] introduced an algorithm that uses the Intel RealSense depth camera and a support vector machine to detect aggressive behavior in pigs; the algorithm exhibited an accuracy of 97.5%. As noted in previous studies, employing 3D camera data under suitable conditions can result in high performance. However, 3D cameras are susceptible to direct sunlight, which hinders their outdoor use. Additionally, their range is limited as they can only capture data effectively within a specific range [13]. Moreover, 3D cameras have lower resolutions and cost more than 2D RGB cameras, rendering their usage on livestock farms challenging [14].

Therefore, some studies have analyzed livestock behavior using 2D image data. Zhang et al. [15] proposed an algorithm that detects the behaviors of sows using image data and classifies them into “drinking”, “urination”, and “mounting” using the MobileNet model; the algorithm exhibited a *mean average precision (mAP)* of 93.4%. Additionally, Wang et al. [16] detected the estrous behavior of cows based on 2D image data using the you only look once v5 (YOLOv5) model, which achieved a *mAP* of 94.3%. Such models that detect livestock behavior based on 2D image data have shown good performances. A significant advantage of these methods is that they enable analysis of various behaviors noninvasively [17]. Moreover, 2D cameras have a wider field of view than 3D cameras and

are relatively stable against direct sunlight, rendering them more suitable for applications on livestock farms. Therefore, this study proposes a system that detects cattle behavior using data collected with 2D RGB cameras. The proposed model was implemented using YOLOv7-E6E, an object detection algorithm. Additionally, we employed data augmentation techniques to mitigate the problem of insufficient data. By detecting cattle behavior in real time, this proposed system can significantly assist livestock farmers.

2. Materials and Methods

2.1. Data Acquisition

The data used for training and validating the proposed cattle behavior detection model were collected at the Gangwon State Livestock Research Institute, Hoengseong, Gangwon State, South Korea. The farm comprised two cows and two calves. Three cameras were installed to capture views from different angles and ensure diverse data collection. Employing data captured from various angles can enhance the generalizability of a model more than those captured from a single angle.

Camera A was installed above the feeding area at a 45° angle to capture the overall view of the pen, whereas Camera B was installed on the upper part of the side and primarily focused on the drinking area and calves' room. Finally, Camera C was installed vertically above the feeding area to capture cattle feeding behaviors. Figure 2 shows the overall structure of the pen and examples of the images captured by each camera.

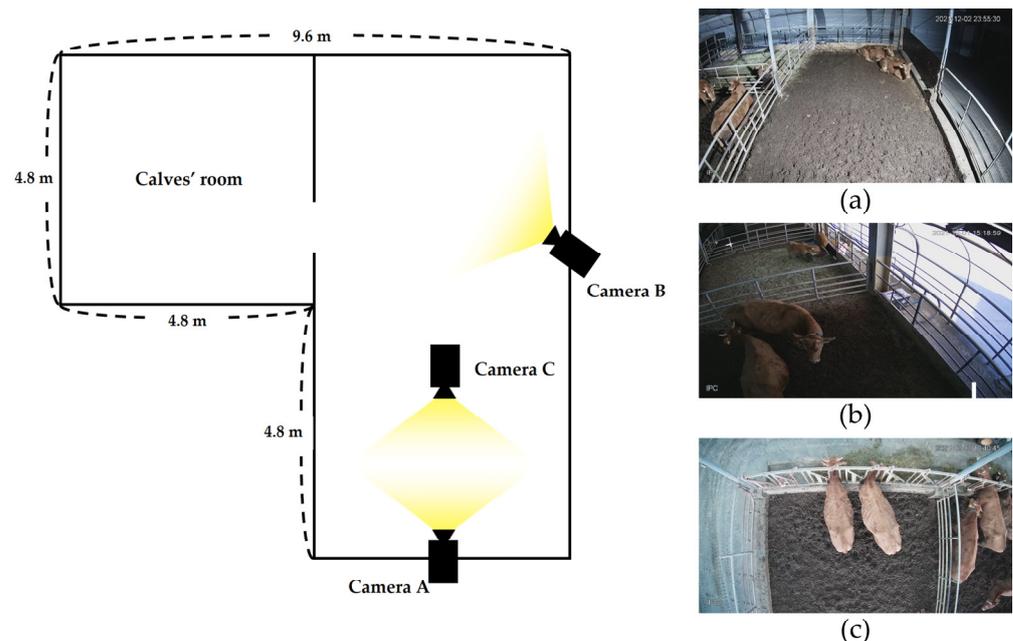


Figure 2. Overall structure of the cattle pen and examples of images captured by (a) camera A, (b) camera B, and (c) camera C.

This study employed network IP cameras (GB-CDX04, GASI) with a focal length in the range of 4.1–16.4 mm and featuring an automatic infrared (IR) function that enables image capture at night. The data captured with this camera were stored in the DAV video format at an HD (1280 × 720 pixels) resolution at 30 fps. The dataset included videos recorded over 264 h, in the period 1–11 December 2021, from different angles using these three cameras.

2.2. Image Extraction

In this study, 792 h of video was converted into images for input into the YOLOv7-E6E model. Image extraction was performed by analyzing the differences in histograms among frames rather than using fixed intervals. Extracting images at regular intervals may

result in repetitive capture of the same image when the cow is stationary or makes minimal movements, and using such similar image data for deep learning model training can induce overfitting. Moreover, the cattle might exhibit unique behaviors between the intervals of the extracted frames, which will not be extracted. Therefore, this study employed an approach that extracts images using the histogram sum differences between frames. Figure 3 shows examples of images extracted using the proposed approach.

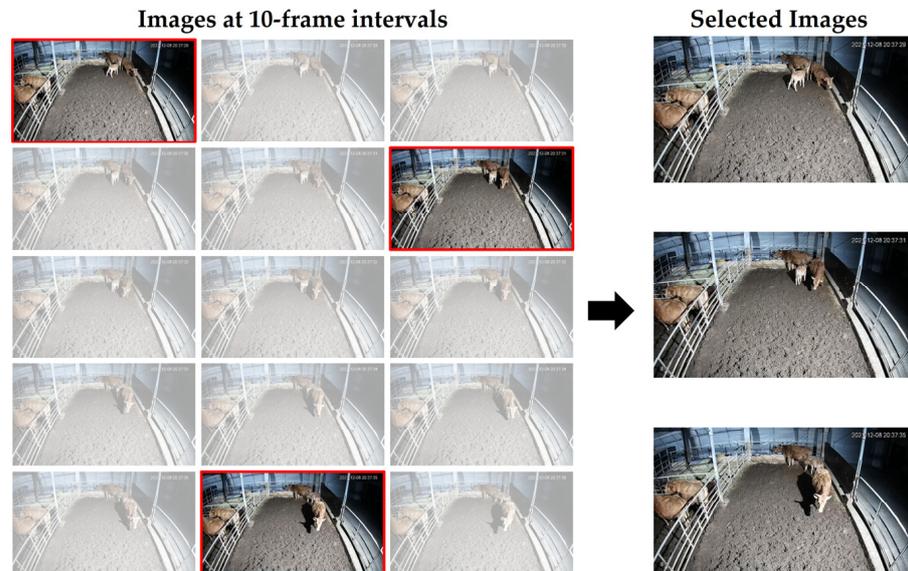


Figure 3. Examples of three consecutive images extracted using differences in histogram sums.

The image extraction method used the following approach to determine frame differences: First, all frames were converted to grayscale, and the absolute gray-level sum of all pixels in each frame was calculated. Subsequently, the first frame was determined as the baseline frame, and the difference between its grayscale sum and that of the second frame was computed. If this difference was $<30,000$, the sum of the differences between the baseline frame and the third frame, which was the next frame, was calculated. This process continued until a frame whose grayscale value differed from that of the baseline frame by $>30,000$ was reached. Thereafter, it was extracted and set as the baseline frame, and the aforementioned process was repeated for all video frames. Figure 4 shows the histograms of the first and second images selected from Figure 3 and the differences between their histograms.

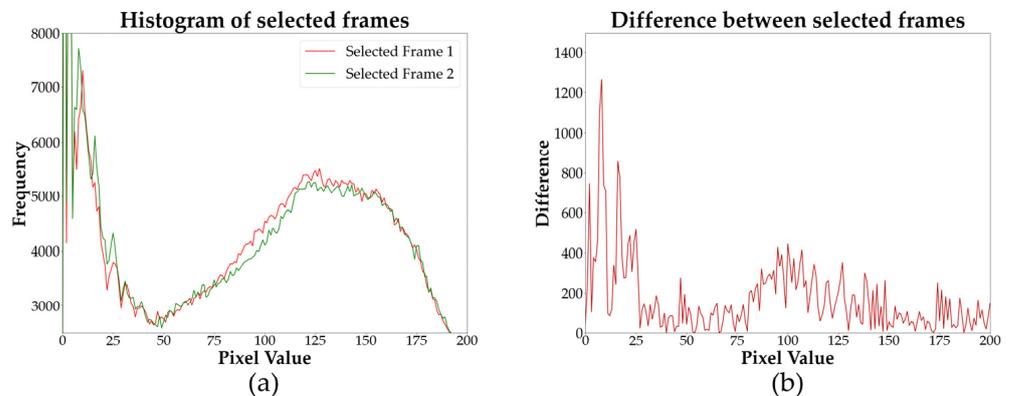


Figure 4. (a) Histogram of the first and second frames extracted in Figure 3; (b) Histogram of the differences between the two frames.

This study analyzed the differences in the gray-level sums to determine a threshold value wherein the difference in the gray-level sums of a frame showing a particular behavior and that showing a different behavior was 30,000. We set this as the threshold of the gray-level sum difference for image extraction.

This method enabled extracting more frames depicting active rather than minimal movement, such as during the rest period. Compared to extraction at constant frame intervals, this approach prevents the accumulation of multiple similar images and effectively extracts frames depicting unique behaviors. Using this method, 18,549 images were extracted at the same resolution as the videos (1280×720 pixels) and stored in the BMP format, which is a lossless and uncompressed image format.

2.3. Dataset Composition

To train the proposed object detection model, areas of the images showing specific behaviors were labeled with bounding boxes and the cattle behaviors were categorized into five classes: resting, communion, feeding, drinking, and eating. These five behavioral classes were further categorized into two groups: independent and interacting. Independent behavior included resting, drinking, and eating, which can be observed independently in cows and calves. Interacting behavior comprised communion and feeding behaviors, which involve interactions between cows and calves. Sample images depicting each behavior class are shown in Figure 5. The behaviors for each class were classified as follows: Resting involved the cattle sitting with their bellies touching the ground. Communion involved the cow's mouth touching the calf's body, or vice versa. Feeding involved the calf placing its head on the cow's teat and drinking milk. Drinking involved passing the entire head over the fence toward the water trough and drinking water. Finally, eating involved placing their heads in the feed trough.

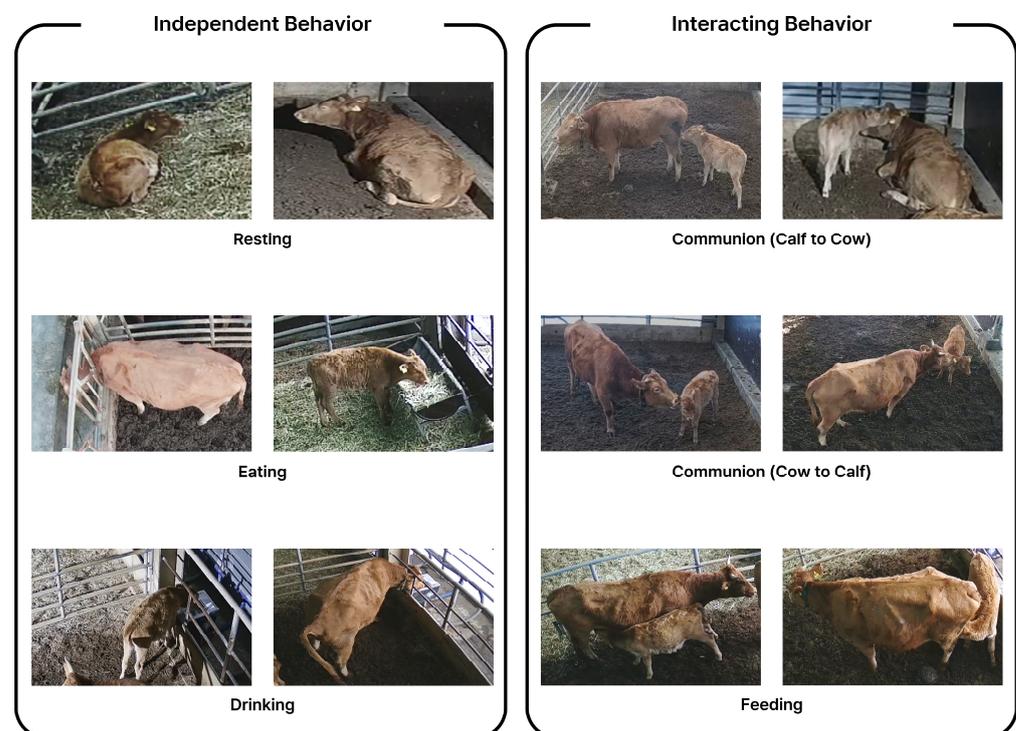


Figure 5. Classification and sample images of each behavior class.

Among the five behavior classes, this study focused on the communion class. Because cattle are social animals, the bond formation between cows and calves is crucial; cattle that lack adequate bonding experience stress. Interactions between cows and calves can strengthen this bond and potentially aid in calming calves [18]. Therefore, bonding behavior is a vital indicator for analyzing cattle behavior.

After labeling, the data were divided into training, validation, and test sets in a ratio of 6:2:2. However, in a deep learning model, employing multiple images from a single video across different datasets can result in overfitting. Therefore, we ensured that the images extracted from a video were included in only one dataset to avoid duplicates among the training, validation, and test sets. Table 1 presents the dataset configurations and Table 2 lists the number of labels for each class.

Table 1. Dataset configurations.

Camera	Train	Validation	Test	Total
Camera A	3723	1235	1230	6188
Camera B	3711	1237	1236	6184
Camera C	3720	1234	1223	6177
Total	11,154	3706	3689	18,549

Table 2. Number of labels for each class.

Class	Train	Validation	Test	Total
Resting	10,567	4396	3218	18,181
Communion	1186	114	846	2146
Feeding	851	173	416	1440
Drinking	299	93	155	547
Eating	5764	2060	2797	10,591
Total	18,667	6836	7432	32,905

2.4. Data Augmentation

To ensure effective training of a deep learning model, high-quality data in large quantities are essential. However, the collection and labeling of such large datasets require significant time and labor. Therefore, this study used data augmentation techniques such as AutoAugment and GridMask.

AutoAugment employs a reinforcement learning algorithm trained on a specific dataset to determine the most effective augmentation policy [19]. There exist 16 types of image augmentation methods: shearX (Y), translateX (Y), rotate, AutoContrast, invert, equalize, solarize, posterize, contrast, color, brightness, sharpness, cutout, and sample pairing. AutoAugment produces 25 pairs of techniques by combining two techniques from this list. The 25 pairs of generated techniques are applied with varying probabilities and intensity levels to introduce diverse modifications to the images, thereby enabling the model to become more robust to variations. This study compared the performances of the augmentation policies trained on the Cifar-10 and ImageNet datasets [20,21]. After data augmentation, the number of training images was increased by $25\times$, from 11,154 to 278,850.

Data augmentation methods such as cutout, hide-and-peek, and GridMask based on information removal encourage the object detection model to focus on learning from a variety of features of the object rather than relying solely on specific characteristics [22–24]. This approach not only improves the model’s generalization performance but also enhances its robustness in diverse environments.

Among the data augmentation methods based on information removal, GridMask uniquely deletes rectangular areas uniformly and masks parts of the image to enhance the diversity of the training data. As a result, the model becomes more robust to various modifications and changes present within the dataset. Furthermore, integrating GridMask introduces randomness and unpredictability into the training phase, serving as a preventive measure against overfitting. It ensures the model remains effective when encountering previously unseen data.

In this study, GridMask was applied to 278,850 images that had been augmented using AutoAugment. GridMask was implemented with a 50% probability, ensuring that the augmentation process introduced a balanced mixture of obscured and unobscured images. The combination of these two augmentation techniques increases the diversity of the dataset and, consequently, the robustness and generalization performance of the object detection model.

2.5. Object Detection Using YOLOv7-E6E

In this study, we used the YOLOv7-E6E model for real-time recognition and classification of cattle behavior [25]. The YOLOv7-E6E model is an advanced iteration of the basic YOLOv7 network, designed to enhance real-time object detection performance and reduce inference costs. It employs several techniques, which are described as follows: First, the model adopts the extended efficient layer aggregation network (E-ELAN), which addresses the limitations of ELAN [26]. E-ELAN efficiently controls the gradient paths through the expansion, shuffle, and merge cardinalities. This enhances the training capability of the model even if many computation blocks are stacked.

Next, an auxiliary head generates fine and coarse labels based on the predictions of the lead head. The fine and coarse labels are used to train the lead and auxiliary heads, respectively. Through this bifurcated training approach, the model can achieve a balanced understanding of both detailed and comprehensive features, leading to performance improvements.

Additionally, training the Aux head for coarse labels enhances the model’s recall performance. It is particularly useful in behavior analysis, where detecting subtle and diverse patterns of behavioral changes is crucial. Moreover, the YOLOv7-E6E model demonstrates better speed and accuracy compared to other methods. Therefore, this study deemed it suitable for real-time detection of cow behaviors. Figure 6 illustrates the structure of the proposed YOLOv7-E6E model, and Figure 7 shows the flowchart of cow behavior detection using the YOLOv7-E6E model.

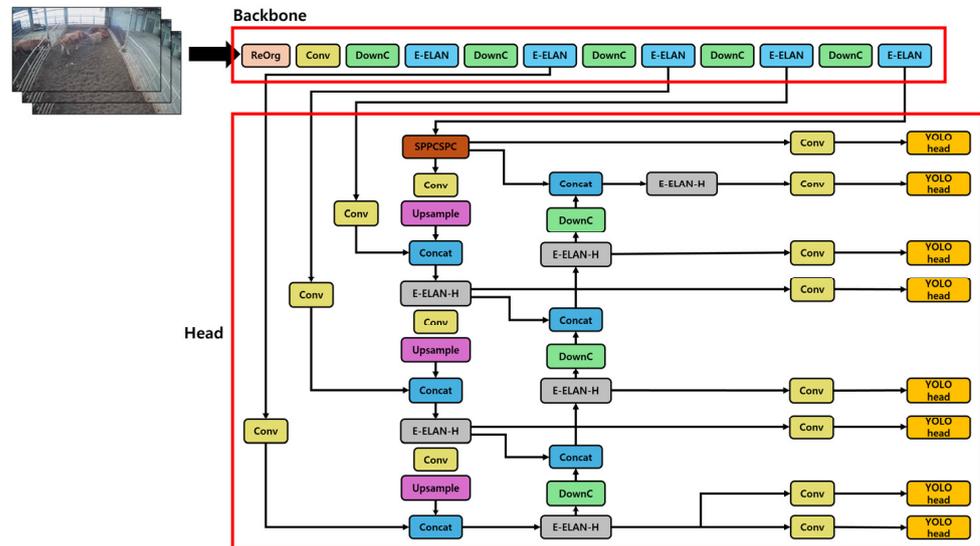


Figure 6. Structure of YOLOv7-E6E.

This study employed a system comprising an Intel i9-10920X CPU @3.50 GHz and NVIDIA GeForce RTX 3090 GPU. The programming environment comprised Windows 10, Python 3.9.12, CUDA 11.8, PyTorch 1.11.0, and Torchvision 0.12.0.

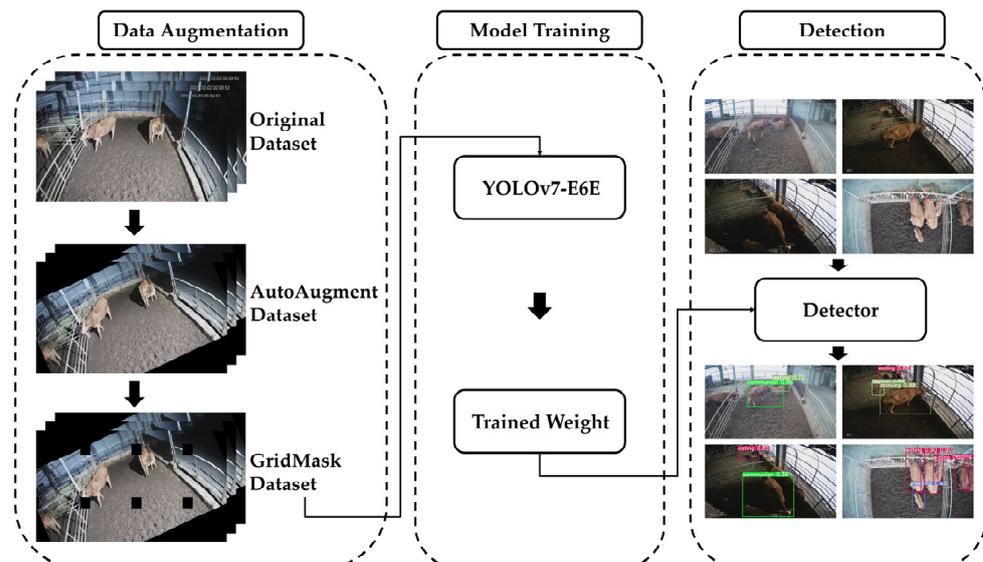


Figure 7. Flowchart of cattle behavior detection using the proposed YOLOv7-E6E model.

2.6. Evaluation Metrics

This study used the *precision*, *recall*, and *mAP* evaluation metrics to evaluate the proposed deep learning model. *Precision* is the ratio of correctly detected objects to the total number of objects identified by a model. *Recall* is the ratio of correctly detected objects by the model among the actual objects. Therefore, a high *precision* value indicates that the model accurately identifies areas containing objects with minimal false detections, whereas a high *recall* value indicates that the model rarely fails to detect the actual behavior of objects. *mAP*, which is commonly used as a performance evaluation metric for object detection models, is the mean of the average *precision* (*AP*) for each class based on an intersection over a union threshold of 0.5. *AP* refers to the area under the *precision–recall* (*P–R*) curve, which illustrates the change in *precision* values under varying *recall* levels.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$AP = \text{Area of Precision – Recall Curve}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

True positive (*TP*) refers to instances wherein the model correctly identifies an object in an image, whereas false positive (*FP*) indicates the incorrect identification of an object that is absent. Finally, false negative (*FN*) indicates that the model fails to detect an object.

3. Results and Discussion

3.1. Original Dataset Results

First, to compare and evaluate the methods used in this study, we analyzed the performance of the model trained on the original dataset; the results are listed in Table 3.

The results for the original image dataset showed that the performance for the communion and feeding classes was relatively lower than that for other classes. This performance degradation is attributed to the characteristics of these classes, as unlike the other classes, these classes involve interactions between cows and calves. The significantly low *recall* rate indicates that the model found it challenging to accurately detect the communion class, which was crucial for this study.

Table 3. Performance of cattle behavior detection for the original dataset.

Behavior	Precision (%)	Recall (%)	mAP (%)
Resting	92.3	86.6	92.6
Communion	84.3	68.7	81.4
Feeding	82.2	67.8	73.7
Drinking	92.5	89.0	96.4
Eating	90.2	95.9	96.4
All	88.3	81.6	88.2

3.2. AutoAugment Dataset Results

As mentioned previously, communion is an important class for bond formation between cows and calves, which is essential for stress reduction among calves. Therefore, AutoAugment with the Cifar-10 and ImageNet policies was applied to the original image dataset to address the relatively low detection performance of the model for this class, and its performance was re-evaluated; the results are listed in Table 4.

Table 4. Performance of the cattle behavior detection model for the dataset augmented using AutoAugment with the Cifar-10 and ImageNet policies.

Augmentation Policy	Behavior	Precision (%)	Recall (%)	mAP (%)
Cifar-10	Resting	93.3	91.7	95.0
	Communion	87.6	81.6	87.0
	Feeding	79.4	71.4	74.5
	Drinking	95.0	99.4	98.9
	Eating	87.5	96.1	97.0
	All	88.6	88.0	90.5
ImageNet	Resting	94.0	88.4	93.5
	Communion	83.7	79.9	87.2
	Feeding	83.7	69.0	78.9
	Drinking	96.8	96.6	98.8
	Eating	88.2	96.4	97.0
	All	89.3	86.1	91.1

The results show that the application of the Cifar-10 and ImageNet policies improved the *precision* for each class by 0.3 and 1.0%, respectively, *recall* by 6.4 and 4.5%, respectively, and *mAP* by 2.3 and 2.9%, respectively. These results indicate that the size of the original image dataset was insufficient to effectively train the model on the object features. Thus, training on the image data generated using these augmentation policies enhanced the detection performance of the model.

Additionally, the ImageNet policy improved the performance more than the Cifar-10 policy. This is because the ImageNet policy is tuned for 1000 classes, which is more than that of Cifar-10, and enabled training the object detection model on a more diverse and robust training set.

3.3. GridMask Data Augmentation Results

The results in Section 3.2 indicate that the ImageNet policy enhances performance more than the Cifar-10 policy. Therefore, GridMask was applied to the images augmented using the ImageNet policy with a 50% probability. The detection performances for training the model on data augmented using AutoAugment and GridMask are presented in Table 5.

Compared with the original dataset, the *precision* decreased by 0.5%, *recall* increased by 9.2%, and *mAP* improved by 4.8%, indicating that GridMask enhanced the generalization performance of the detection model. Additionally, the model could recognize object features using only partial information regarding the object. Consequently, its object detection capability in various environments was improved. The integration of GridMask and

AutoAugment enhanced the training dataset diversity, resulting in notable improvements in the model performance. However, owing to a 6.4% decrease in the *precision* score for the eating class, the *precision* for all classes dropped by 0.5%. The eating class labeled behaviors such as cows putting their heads through the feeding fence or placing a calf's head inside the feed container. Therefore, the labels exhibited two distinct positions and shapes. By applying AutoAugment and GridMask, parts of the feed container and fence were obscured in the image transformations, preventing adequate model training on these behaviors. Therefore, cattle that did not exhibit any specific behavior were incorrectly detected as eating, which decreased the *precision* for the eating class.

Table 5. Performance of the cattle behavior detection model trained on data augmented using AutoAugment and GridMask.

Behavior	Precision (%)	Recall (%)	mAP (%)
Resting	92.4	93.8	95.0
Communion	82.1	86.7	90.1
Feeding	84.5	77.4	83.2
Drinking	96.2	98.8	98.9
Eating	83.8	97.5	97.7
All	87.8	90.8	93.0

Recall and *mAP* are important metrics for livestock farming. Zheng et al. [27] noted that high *recall* and *mAP* scores help effectively detect cattle. Additionally, a high *recall* score for livestock behavior detection minimizes the non-detection rate of behaviors related to diseases or estrus. Thus, it is directly related to livestock health and ultimately affects the productivity of agricultural farms and is a crucial evaluation metric. The proposed detection model showed a 9.2% improvement in *recall* when it was trained on data augmented using AutoAugment and GridMask compared to the original dataset. This improvement is expected to provide valuable information, potentially leading to a substantial increase in labor efficiency and reduced labor costs for farmers.

Figure 8 presents the detection results of the model with both AutoAugment and GridMask applied. Upon reviewing the detection outcomes of the model integrating these two augmentation techniques, it becomes evident that the synergy between AutoAugment and GridMask significantly enhances the model's capability to detect and classify cattle behaviors accurately. The combined use of these techniques minimizes false detections and strengthens the recognition of diverse behaviors across various environmental conditions, resulting in clear and well-defined object detections. It not only demonstrates the efficiency of the combined augmentation strategies but also highlights the potential of our proposed model to reliably capture subtle cattle behaviors, offering a promising tool for applications in precision livestock farming.

Pavlovic et al. [8] utilized tail-mounted accelerometers to classify cattle behavior, achieving a *sensitivity (recall)* and *precision* of over 86.7%. Williams et al. [9] classified cattle behavior using accelerometer necklaces, achieving an average F1 score of 0.82. However, our study significantly differs from these approaches as we only used 2D RGB camera data and did not attach invasive sensors to the cattle. This approach not only simplifies the data collection process but also potentially reduces stress on the cattle. This study achieved a 93% *mAP*, demonstrating the efficiency of using 2D RGB camera-based observations for high-accuracy cattle behavior detection.

This study proposed a cattle behavior detection model based on YOLOv7-E6E. Cattle behavior data were collected using 2D RGB cameras, and the model performance was enhanced by training it on the original data augmented through AutoAugment. Additionally, training on data augmented through GridMask decreased *precision* by 0.5% compared to the original dataset but increased the *recall* and *mAP* by 9.2% and 4.8%, respectively.

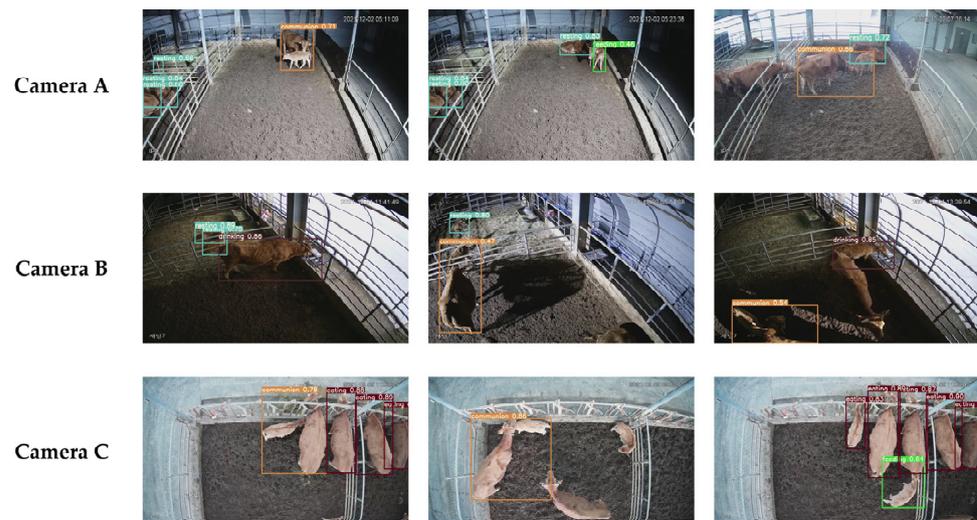


Figure 8. The detection results of proposed model.

4. Conclusions

The livestock sector faces challenges due to a decreased labor force, increasing workloads, and heightened stress for farmers [28]. In response, this study proposes a cattle behavior detection model utilizing 2D RGB cameras to automate and simplify monitoring processes within livestock farms. In this study, we enhanced model performance by 9.2% for *recall* and 4.8% for *mAP* using AutoAugment and GridMask compared to the original datasets. This automation can reduce the workload on farmers by providing real-time insights into cattle behavior and contributing to the early detection of health issues. It has the potential to significantly cut veterinary costs and reduce losses from livestock health problems. When estrus detection in cattle is considered, they exhibit increased activity, decreased resting, and reduced feed intake, which facilitates estrus detection through behavioral analysis [29]. Employing PLF technology can yield up to EUR 2729 in annual economic benefits per farm over non-PLF methods [30]. Thus, this research could contribute to sustainable livestock farming and profitability through reduced labor costs and increased livestock production.

This study has a limitation in that the *precision* slightly decreased in the model trained on data augmented using AutoAugment and GridMask compared to that trained on the original data. Therefore, future studies should use images enhanced with AutoAugment and GridMask after filtering. Additionally, labels for objects whose features have disappeared or are heavily modified because of augmentation policies can be problematic. Removing these labels using a model trained only on the original data might resolve the performance decline. In addition, this study used data collected from a single farm with the same recording equipment, which could have resulted in model bias toward this specific farm, and its performance might not be adequately validated using data from other farms. To address these issues, we are installing different cameras, not limited to those used in this study, in various barns to collect data. In reviewing various studies on cattle behavior detection, Fuentes et al. [31] employed the YOLOv3 model to detect 15 classes of cow behavior in 1920×1080 resolution images, achieving a *mAP* of 78.8%. Similarly, Uchino and Ohwada [32] achieved a *mAP* of 91.5% utilizing the YOLOv5-L model for four classes at a resolution of 3840×2160 . Our study has shown similar or higher performance despite utilizing lower resolutions. Thus, future experiments incorporating diverse angles, barn environments, and resolutions could mitigate model bias and enhance the model's generalization performance.

In this study, our methodology utilized the sum of gray levels from the entire image. Recognizing the potential for further refinement, we consider exploring a more targeted approach for future research. Specifically, we aim to investigate the application of summing

gray levels within designated regions of interest. This is particularly beneficial in scenarios where vital information is compromised by background noise or other distractive elements. Focusing on regions could alleviate such challenges, facilitating a more accurate feature extraction of the target object. This strategy holds promise in domains where the significance of certain image areas is paramount. For example, in precision livestock farming (PLF), zeroing in on segments of aerial images that capture active livestock could enhance model detection accuracy. As we progress, our goal is to refine this approach for image analysis and streamline its application, making it versatile enough for broader use across various fields. This evolution will involve developing and meticulously tuning the method to automate the process for enhanced efficiency and applicability.

Author Contributions: Conceptualization, H.-c.C. and J.S.K.; methodology, H.-s.S., T.-k.K. and H.-c.C.; software, H.-s.S. and T.-k.K.; validation, H.-s.S. and T.-k.K.; formal analysis, H.-s.S. and T.-k.K.; investigation, H.-s.S. and J.S.K.; resources, T.-k.K.; data curation, H.-s.S., T.-k.K., C.-w.L. and C.-s.C.; writing—original draft preparation, H.-s.S. and T.-k.K.; writing—review and editing, H.-c.C.; visualization, H.-s.S. and T.-k.K.; supervision, H.-c.C.; project administration, H.-c.C.; funding acquisition, H.-c.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R111A3053872). This research was supported by the “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2022RIS-005). This research was supported by Rural Development Administration, Republic of Korea (No. 00260110).

Institutional Review Board Statement: The animal study was approved by the Gangwon State Livestock Research Institute in South Korea on 21 December 2023, with approval number 2023-6.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Clonan, A.; Roberts, K.E.; Holdsworth, M. Socioeconomic and demographic drivers of red and processed meat consumption: Implications for health and environmental sustainability. *Proc. Nutr. Soc.* **2016**, *75*, 367–373. [[CrossRef](#)] [[PubMed](#)]
2. OECD. *Meat Consumption (Indicator)*; OECD: Paris, France, 2024. [[CrossRef](#)]
3. Werkheiser, I. Technology and responsibility: A discussion of underexamined risks and concerns in precision livestock farming. *Anim. Front.* **2020**, *10*, 51–57. [[CrossRef](#)] [[PubMed](#)]
4. Berckmans, D. General introduction to precision livestock farming. *Anim. Front.* **2017**, *7*, 6–11. [[CrossRef](#)]
5. Da Silva Santos, A.; De Medeiros VW, C.; Gonçalves, G.E. Monitoring and classification of cattle behavior: A survey. *Smart Agric. Technol.* **2023**, *3*, 100091. [[CrossRef](#)]
6. Garcia, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodriguez, P. A systematic literature review on the use of machine learning in precision livestock farming. *Comput. Electron. Agric.* **2020**, *179*, 105826. [[CrossRef](#)]
7. Schillings, J.; Bennett, R.; Rose, D.C. Exploring the potential of precision livestock farming technologies to help address farm animal welfare. *Front. Anim. Sci.* **2021**, *2*, 639678. [[CrossRef](#)]
8. Pavlovic, D.; Davison, C.; Hamilton, A.; Marko, O.; Atkinson, R.; Michie, C.; Crnojević, V.; Andonovic, I.; Bellekens, X.; Tachtatzis, C. Classification of cattle behaviours using neck-mounted accelerometer-equipped collars and convolutional neural networks. *Sensors* **2021**, *21*, 4050. [[CrossRef](#)] [[PubMed](#)]
9. Williams, M.; Lai, S.Z. Classification of dairy cow excretory events using a tail-mounted accelerometer. *Comput. Electron. Agric.* **2022**, *199*, 107187. [[CrossRef](#)]
10. Achour, B.; Belkadi, M.; Filali, I.; Laghrouche, M.; Lahdir, M. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on Convolutional Neural Networks (CNN). *Biosyst. Eng.* **2020**, *198*, 31–49. [[CrossRef](#)]
11. Wu, D.; Wu, Q.; Yin, X.; Jiang, B.; Wang, H.; He, D.; Song, H. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosyst. Eng.* **2020**, *189*, 150–163. [[CrossRef](#)]
12. Chen, C.; Zhu, W.; Liu, D.; Steibel, J.; Siegford, J.; Wurtz, K.; Han, J.; Norton, T. Detection of aggressive behaviours in pigs using a RealSence depth sensor. *Comput. Electron. Agric.* **2019**, *166*, 105003. [[CrossRef](#)]
13. Ruchay, A.; Kober, V.; Dorofeev, K.; Kolpakov, V.; Miroshnikov, S. Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery. *Comput. Electron. Agric.* **2020**, *179*, 105821. [[CrossRef](#)]

14. Wurtz, K.; Camerlink, I.; D'Eath, R.B.; Fernández, A.P.; Norton, T.; Steibel, J.; Siegford, J. Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review. *PLoS ONE* **2019**, *14*, e0226669. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, Y.; Cai, J.; Xiao, D.; Li, Z.; Xiong, B. Real-time sow behavior detection based on deep learning. *Comput. Electron. Agric.* **2019**, *163*, 104884. [[CrossRef](#)]
16. Wang, R.; Gao, Z.; Li, Q.; Zhao, C.; Gao, R.; Zhang, H.; Li, S.; Feng, L. Detection method of cow estrus behavior in natural scenes based on improved YOLOv5. *Agriculture* **2022**, *12*, 1339. [[CrossRef](#)]
17. Norton, T.; Chen, C.; Larsen ML, V.; Berckmans, D. Precision livestock farming: Building 'digital representations' to bring the animals closer to the farmer. *Animal* **2019**, *13*, 3009–3017. [[CrossRef](#)]
18. Cooke, S. The Ethics of Touch and the Importance of Nonhuman Relationships in Animal Agriculture. *J. Agric. Environ. Ethics* **2021**, *34*, 12. [[CrossRef](#)]
19. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. Autoaugment: Learning augmentation strategies from data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 113–123.
20. Krizhevsky, A.; Hinton, G. *Learning Multiple Layers of Features from Tiny Images*; Technical Report; University of Toronto: Toronto, ON, Canada, 2009.
21. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
22. DeVries, T.; Taylor, G.W. Improved regularization of convolutional neural networks with cutout. *arXiv* **2017**, arXiv:1708.04552.
23. Singh, K.K.; Yu, H.; Sarmasi, A.; Pradeep, G.; Lee, Y.J. Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. *arXiv* **2018**, arXiv:1811.02545.
24. Chen, P.; Liu, S.; Zhao, H.; Jia, J. Gridmask data augmentation. *arXiv* **2020**, arXiv:2001.04086.
25. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
26. Wang, C.Y.; Liao HY, M.; Yeh, I.H. Designing network design strategies through gradient path analysis. *arXiv* **2022**, arXiv:2211.04800.
27. Zheng, Z.; Li, J.; Qin, L. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Comput. Electron. Agric.* **2023**, *209*, 107857. [[CrossRef](#)]
28. Duval, J.; Cournut, S.; Hostiou, N. Livestock farmers' working conditions in agroecological farming systems. A review. *Agron. Sustain. Dev.* **2021**, *41*, 22. [[CrossRef](#)]
29. Aquilani, C.; Confessore, A.; Bozzi, R.; Sirtori, F.; Pugliese, C. Precision Livestock Farming technologies in pasture-based livestock systems. *Animal* **2022**, *16*, 100429. [[CrossRef](#)] [[PubMed](#)]
30. CORDIS. Final Report Summary—EU-PLF (Bright Farm by Precision Livestock Farming). 2016. Available online: <https://cordis.europa.eu/project/id/311825/reporting> (accessed on 7 April 2024).
31. Fuentes, A.; Yoon, S.; Park, J.; Park, D.S. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Comput. Electron. Agric.* **2020**, *177*, 105627. [[CrossRef](#)]
32. Uchino, T.; Ohwada, H. Individual identification model and method for estimating social rank among herd of dairy cows using YOLOv5. In Proceedings of the 2021 IEEE 20th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), Banff, AB, Canada, 29–31 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 235–241.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.