



Assessing optimal frequency for image acquisition in computer vision systems developed to monitor feeding behavior of group-housed Holstein heifers

T. Bresolin, , R. Ferreira, , F. Reyes, , J. Van Os, , and J. R. R. Dórea*

Department of Animal and Dairy Sciences, University of Wisconsin, Madison 53706

ABSTRACT

Computer vision systems have emerged as a potential tool to monitor the behavior of livestock animals. Such high-throughput systems can generate massive redundant data sets for training and inference, which can lead to higher computational and economic costs. The objectives of this study were (1) to develop a computer vision system to individually monitor detailed feeding behaviors of group-housed dairy heifers, and (2) to determine the optimal frequency of image acquisition to perform inference with minimal effect on feeding behavior prediction quality. Eight Holstein heifers (96 ± 6 d old) were housed in a group and a total of 25,214 images (1 image every second) were acquired using 1 RGB camera. A total of 2,209 images were selected and each animal in the image was labeled with its respective identification (1–8). The label was annotated only on animals that were at the feed bunk (head through the feed rail). From the labeled images, 1,392 were randomly selected to train a deep learning algorithm for object detection with YOLOv3 (“You Only Look Once” version 3) and 154 images were used for validation. An independent data set (testing set = 663 out of the 2,209 images) was used to test the algorithm. The average accuracy for identifying individual animals in the testing set was 96.0%, and for each individual heifer from 1 to 8 the accuracy was 99.2, 99.6, 99.2, 99.6, 99.6, 99.2, 99.4, and 99.6%, respectively. After identifying the animals at the feed bunk, we computed the following feeding behavior parameters: number of visits (NV), mean visit duration (MVD), mean interval between visits (MIBV), and feeding time (FT) for each heifer using a data set composed by 8,883 sequential images (1 image every second) from 4 time points. The coefficient of determination (R^2) was 0.39, 0.78, 0.48, and 0.99, and the root mean square error (RMSE) were 12.3 (count), 0.78, 0.63, and 0.31 min for NV, MVD,

MIBV, and FT, respectively, considering 1 image every second. When we moved from 1 image per second to 1 image every 5 (MIBV) or 10 (NV, MDV, and FT) s, the R^2 observed were 0.55 (NV), 0.74 (MVD), 0.70 (MIBV), and 0.99 (FT); and the RMSE were 2.27 (NV, count), 0.38 min (MVD), 0.22 min (MIBV), and 0.44 min (FT). Our results indicate that computer vision systems can be used to individually identify group-housed Holstein heifers (overall accuracy = 99.4%). Based on individual identification, feeding behavior such as MVD, MIBV, and FT can be monitored with reasonable accuracy and precision. Regardless of the frequency for optimal image acquisition, our results suggested that longer time intervals of image acquisition would reduce data collecting and model inference while maintaining adequate predictive performance. However, we did not find an optimal time interval for all feeding behavior; instead, the optimal frequency of image acquisition is phenotype-specific. Overall, the best R^2 and RMSE for NV, MDV, and FT were achieved using 1 image every 10 s, and for MIBV it was achieved using 1 image every 5 s, and in both cases model inference and data storage could be drastically reduced.

Key words: deep learning, feeding time, image labeling, machine learning, prediction accuracy, visit duration

INTRODUCTION

Feeding behavior has received increased attention for its association with productivity and efficiency in dairy cattle, which might reduce the environmental footprint (von Keyserlingk and Weary, 2010). Feeding behavior of a dairy cow can be measured by feeding time (FT), meal duration, meal frequency, and feeding rate (Nielsen, 1999). Increasing the meal duration, for example, facilitates chewing, reduces feed particles size, and increases digestibility (Aikman et al., 2008). Greater FT increases the production of saliva, which decreases acidity in the rumen (Beauchemin et al., 2008), and by reducing the feeding rate, the risk of metabolic disorders is also reduced (Shaver, 1997). In

Received March 29, 2022.

Accepted August 2, 2022.

*Corresponding author: joao.dorea@wisc.edu

addition, feeding behavior has been considered an important indicator of feeding preference (Nielsen et al., 2000), milk composition (Gregorini et al., 2013), early detection of estrus (Cairo et al., 2020), and detection of respiratory diseases (Duthie et al., 2021) in dairy cattle. Monitoring detailed feeding behavior is a crucial process; however, it is unfeasible to be performed systematically in large-scale operations (Britt et al., 2018). Additionally, the assessment of feeding behavior patterns based on human observation is time consuming, prone to error, and requires a subject-matter expert (Nielsen, 2013; Mattachini et al., 2016). Therefore, there has been increasing interest for new technologies capable of assessing individual animal feeding behavior in real-time, which could save time and resources, and increase the herd health and profitability of the dairy farms (Nielsen, 2013; Mattachini et al., 2016; Duthie et al., 2021).

Precision livestock technologies, such as wearable sensors, have been proposed as an alternative to collect difficult-to-measure information in cattle (Williams et al., 2019; Cairo et al., 2020; Coelho Ribeiro et al., 2021). Wearable sensors, however, are usually phenotype-specific and require 1 device per animal. Conversely, computer vision has become a potential technology for collecting individual feeding behavior data, as few devices can monitor a group of several animals, and the images collected are not limited to a specific phenotype (i.e., multiple phenotypes could be generated using the same hardware infrastructure). Thus, the use of computer vision systems has emerged as a powerful tool to monitor animal behavior in different species, such as swine (Brown-Brandl et al., 2013; Yang et al., 2020), chickens (Aydin et al., 2010; Neves et al., 2015), goats (Jiang et al., 2020; Rao et al., 2020), and cattle (Porto et al., 2013, 2015). Despite its importance, studies evaluating detailed individual feeding behavior (e.g., number of feeding visits, visit duration, and FT) of dairy heifers using computer vision techniques, to our knowledge, have not yet been described in the literature.

Although some research studies have demonstrated the feasibility of computer vision systems to generate accurate phenotypes, the assessment of optimal frequency of image acquisition in a computer vision system for livestock operations has not been studied yet. This is an extremely important factor given the amount of data generated by such vision systems, usually with high degree of redundancy. Such high-dimensional data requires large computational infrastructure for model training and inference, and data storage (Ninga et al., 2020). Feeding behavior, for example, is characterized by activities lasting minutes or even hours (Cairo et al., 2020); therefore, model inference performed every second may not bring any additional benefit as compared

with a longer inference interval. The development of computer vision systems that optimizes model inference, training, or both is crucial to reduce the costs associated with technology implementation and data processing in farm settings. In this study, we aimed (1) to evaluate the potential of a computer vision system based on a deep learning algorithm called “You Only Look Once” version 3 (YOLOv3) to predict detailed feeding behavior, including the number of visits (NV), mean visit duration (MVD), the mean interval between visits (MIBV), and FT of group-housed dairy heifers based on animal detection and identification, and (2) to assess the optimal frequency of image acquisition to perform inference with minimal effect on the quality of feeding behavior prediction.

MATERIALS AND METHODS

The computer vision system developed in this study to individually identify and monitor detailed feeding behavior follows 4 main steps: (1) image acquisition from 8 group-housed heifers was acquired through video recording; (2) image selection for labeling with a bounding box if the heifer was with head through the feed rail, which is associated with eating behavior; (3) algorithm training to detect and identify each heifer was performed using labeled images; and (4) algorithm deployment on a set of sequential images sampled from 4-time points to predict feeding behavior. The multi-step procedure designed for this study is described in the next 2 sections. All animal protocols were approved by the University of Wisconsin—Madison College of Agricultural and Life Sciences Animal Care and Use Committee.

Animals and Image Acquisition

Eight Holstein heifers on average 96 ± 6 d old, group-housed at the Marshfield Agricultural Research Station of the University of Wisconsin—Madison, were used to develop a feeding behavior monitoring algorithm. Heifers were fed grain (0730 and 1515 h) and unchopped grass hay (0830 and 1600 h) twice daily, whereas water was offered ad-libitum. The animals were housed in a single pen with dimensions 5 m \times 5 m, which represent a total of 25 square footage. A 3 MP ProHDDome IP Fixed Outdoor Camera (Amcrest), placed in the barn ceiling at 3.5 m high, pointing straight toward the feed bunk, was used for video recording. The camera height and position were set to maximize the feeding area view. Seven hours of video (from 1330 to 2030 h) encoded at 20 frames per second, with a frame resolution of $1,080 \times 1,920$ pixels, and 8-bit RGB color space was streamed. A total of 25,214 images were extracted



Figure 1. Frame example from the streaming video projected onto the feed bunk area before (left) and after (right) labeling.

from the video by acquiring 1 frame every second. Five time points (T1 = 13:32:50 to 14:09:38; T2 = 14:40:00 to 15:18:28; T3 = 16:20:12 to 17:08:40; T4 = 17:35:32 to 18:05:40; T5 = 18:28:02 to 18:58:57) matching when heifers were with the head through the feed rail, which is associated with eating behavior, were identified from the total set of images, resulting in a data set with 11,092 sequential images, 1 image every second. From the first time point, 2,209 images (Data 1) were manually labeled using the Image Labeler tool from Computer Vision Toolbox 9.1 (Matlab R2019b). For each of the 2,209 images, heifers had their head through the feed rail received a label with their respective identification (from 1 to 8), resulting in total of 7,457 labels. Considering that a sequence of images was selected to be labeled, heifers at the feed bunk were not constant,

which resulted in an imbalanced labeled image data set. Therefore, the algorithm was trained to identify when the heifers had their head through the feed rail, which is associated with feeding behavior (see more details in Animal Identification section). The remaining 8,883 images (Data 2) from time points 2 to 5 were used to predict the feeding behavior described in the Feeding Behavior section. An example of a snapshot frame before and after labeling is presented in Figure 1 and the number of labels by heifer is shown in Figure 2.

Deep Learning Approach

A deep learning object detection approach, called “You Only Look Once” version 3, YOLOv3 (Redmon and Farhadi, 2018), was used to develop an algorithm

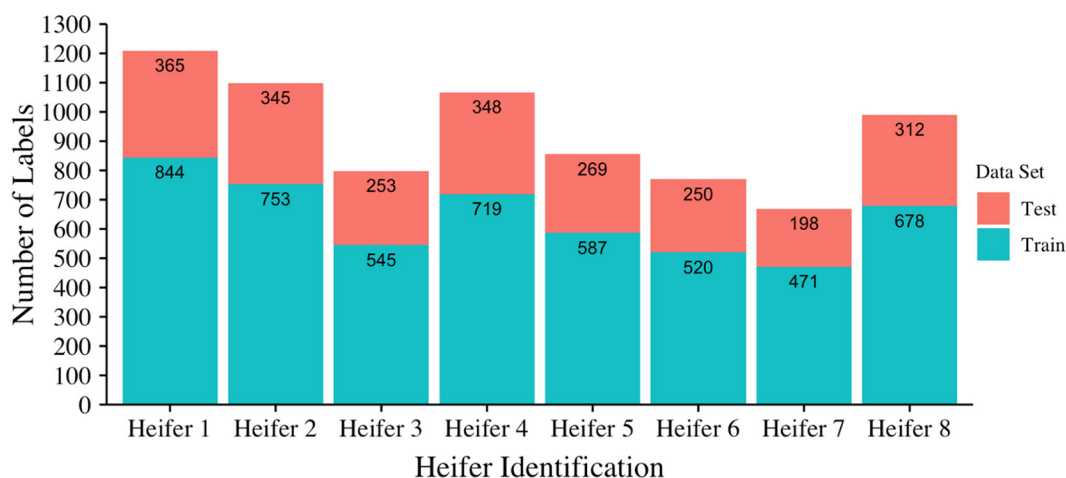


Figure 2. Label distribution for each individual dairy heifer in the training and testing image data sets.

for detection and feeding behavior monitoring. The YOLO system applies a single forward pass neural network to the whole image and predicts bounding box locations (detection) and their respective class probabilities (classification). The YOLOv3 network structure is based on Darknet-53, for feature extraction, each followed by a batch normalization layer (i.e., technique used to standardize the inputs to a layer for each mini-batch) and Leaky Rectified Linear Unit activation function (i.e., linear function that outputs the input directly if it is positive; otherwise, it will output zero). For detection, YOLOv3 uses 53 additional convolutional layers resulting in a network with a total of 106 layers (fully convolutional neural network).

Given an image as input to the YOLO algorithm, it divides the image into an $S \times S$ grid of cells, with S as a parameter defined by the user. If the center of a target object falls into a grid cell, that cell is responsible for detecting the object, in our case an individual heifer. For each grid cell, the network outputs several bounding boxes, class probabilities, and a confidence score. Four values for each bounding box are predicted: the x and y coordinates of the center of the box, and its width and height. The class probabilities are the probabilities of the enclosed object being each of the target classes (heifers 1 to 8). The confidence score is the probability that an object (e.g., heifer 1) is within the predicted bounding box, which is the intersection over union between the predicted and ground truth bounding box. Most of the bounding boxes generated are eliminated, using a technique called non-maximum suppression, either because their confidence scores are below a certain threshold, or they are enclosing the same object as another bounding box with a higher confidence score. The YOLOv3 network works very similarly to the original YOLO algorithm, but with the main contribution that it makes detection at 3 different scales, using 3 different grid sizes at the same time. The algorithm up samples the feature maps extracted from the first 53 convolutional layers by factors of 2 and 4, and then further applies convolutions and 1×1 detection kernels on top of them. This results in 3 distinct detection feature maps per image, each having dimensions 32, 16, and 8 times smaller than the input image, representing the predictions at those 3 different scales. Such design allows for better results when analyzing different sized objects in a single image, which is particularly useful for detecting small objects (Redmon and Farhadi, 2018).

Before being fed into the YOLOv3 network, image pixels were scaled by a factor of $1/255$ to fit in a range from 0 to 1 and were resized from $1,080 \times 1,920 \times 3$ (height, width, and color channels) to $416 \times 416 \times 3$ without cropping, which are the same dimensions used

in the experiments published by the YOLOv3 authors (Redmon and Farhadi, 2018). The resized images were fed into the network in batches of size 32. Weights extracted from a network trained using the MSCOCO data set (Lin et al., 2014) were used as a starting point (i.e., a process called transfer learning) to retrain our YOLOv3 network to identify heifers 1 through 8 while they were at the feed bunk (head through the feed rail). Transfer learning was employed by freezing the first 103 layers for 51 epochs using our image data set, along with label information for each heifer, while the last 3 layers were retrained. Transfer learning is a powerful technique, especially in large data sets, that decreases the computation and time resources required to develop a neural network algorithm. The algorithm was retrained using an Adam optimizer with a starting learning rate of 0.001. On a second training step, all the layers were unfrozen, and the network was trained for 25 epochs with a batch size of 4. We implemented YOLOv3 in Python version 3.6 (<https://www.python.org/>) using Keras version 2.2 (Chollet et al., 2015) and Tensorflow version 1.5 as backend (Abadi et al., 2015; available at <https://github.com/AntonMu/TrainYourOwnYOLO>). All analyses were performed using the computer resources of the University of Wisconsin-Madison Center for High Throughput Computing.

Animal Identification

For animal detection and identification, the first 1,546 (out of 2,209) sequential images from Data 1 were used to retrain YOLOv3, with 1,392 images randomly selected to train, and 154 used to validate the model. The algorithm reached an average loss function value of 18.37 and 17.55 for the training and validation set, respectively, using a learning rate of 10^{-8} . The final algorithm testing was performed on the remaining 663 sequential images from Data 1 that were not exposed to the network during training. To evaluate the algorithm performance for identifying each heifer (1–8) in the testing set, accuracy, sensitivity, specificity, positive predicted value, and negative predicted value metrics were calculated using the following equations: accuracy = $(TP + TN)/(TP + TN + FP + FN)$, sensitivity = $TP/(TP + FN)$, specificity = $TN/(TN + FP)$, positive predicted value = $TP/(TP + FP)$, and negative predicted value = $TN/(TN + FN)$, where TP, TN, FP, and FN represent true positive, true negative, false positive and false negative, respectively. These parameters are interpreted as follows: TP represents the number of times the algorithm correctly identified that heifer (“heifer 1”); TN is the number of times that the algorithm correctly identifies a heifer who is not “heifer 1” as not being “heifer 1”; FP is the number of

times another heifer (e.g., 2–8) was incorrectly identified as “heifer 1”; and FN is the number of times the algorithm incorrectly predicted that “heifer 1” was one of the other heifers.

Feeding Behavior

After training and testing the algorithm performance to identify heifers at the feed bunk, the trained algorithm was applied on the 8,883 sequential images (1 image every second) from Data 2 (see Animals and Image Acquisition section for more details). Based on the identification of each heifer at the feed bunk, we calculated the individual NV, MVD, MIBV, and FT behaviors for each time period, described in Animals and Image Acquisition section, and defined as follows: (1) NV was the total number of times a heifer put its head through the feed rail; (2) MVD was the time between a heifer putting its head through the feed rail and then backing out, averaged across all visits; (3) MIBV was the time between each visit at the feed bunk, averaged across all visits; and (4) FT was the sum of the duration of all visits to the feed bunk (i.e., head through the feed rail). The 8,883 sequential images were manually label by the same person if the heifers were with the head through the feed rail. In both predicted and observed labels, a new visit was defined every time a heifer was identified with her head through the feed rail until she backed out of the feed bunk. The time duration of each behavior was computed based on the number of images, which were acquired every second. To evaluate the predictive quality of the algorithm, we calculated the coefficient of determination (R^2) and the root mean square error (RMSE).

Frequency of Image Acquisition

Using the 8,883 sequential images from Data 2 (see Animals and Image Acquisition section for more details), we created 8 scenarios to investigate the optimal frequency of image acquisition to predict specific feeding behaviors (NV, MVD, MIBV, and FT). Each scenario was based on images sampled every 1 (baseline), 5, 10, 20, 30, 60, 90, and 120 s. The total time for each behavioral activity was computed based on the animal identification at the feed bunk and time interval with which images were sampled. For example, if 1 animal was identified at the feed bunk in 10 sequential images obtained every second, then total FT was 10 s. Thus, with the proposed time sample intervals, we could investigate the longest time interval in which feeding behavior activity can be predicted without compromising model precision and accuracy. We used R^2 and RMSE

between observed and predicted feeding behaviors to evaluate the optimal frequency for image acquisition.

RESULTS AND DISCUSSION

Animal Identification

A computer vision system was developed in this study using YOLOv3 deep learning approach to individually identify group-housed heifers while they were at the feed bunk (head through the feed rail). The identification algorithm predicted a total of 2,383 different labels across 8 classes (heifers 1 to 8) when applied on the testing set. From the total of predicted labels, 1,961 (83.9%) were predicted with a confidence score higher than 0.90, 106 (4.5%) with confidence score between 0.80 and 0.90, 95 (4.1%) with confidence score between 0.70 and 0.80, and 176 (7.5%) were predicted with confidence score in a range from 0.25 to 0.69. A confusion matrix for the labels predicted using the testing set is presented in Supplemental Table S1 (<http://dx.doi.org/10.17632/4gf2tcx8n4.1>; Bresolin, 2022). The algorithm correctly classified 2,286 (95.9%) labels, whereas 9 (0.4%) labels were incorrectly classified, 45 (1.9%) ground-truth labels were not predicted, and 43 (1.8%) predicted labels did not have ground-truth labels. The overall accuracy for predicting a heifer's identity in this study was 96%, which is in the range (84.2 to 96.87%) reported in the literature using the color pattern, collar number, or muzzle point in dairy cows (Kumar et al., 2018; Okura et al., 2019; Bezen et al., 2020). Class imbalance is a problem often found with real-world data sets that can negatively affect the algorithm performance (Valova et al., 2020). Although the number of labels used in both training and testing sets across heifers were different in this study (Figure 2), the imbalance data set did not affect the accuracy for identifying each heifer, as shown in Table 1. One possible explanation for this result is that sufficient feature representation of the heifers was captured by the algorithm in the training data set used in this study, considering that the camera was stationary, and the animal posture did not present drastic changes due to the headlock constraints. This fact is a positive aspect of the proposed computer vision system because cameras positioned to capture animals' rear view at the feed bunk could leverage small changes in body posture, and consequently minimize the need of large feature representations per object to maintain adequate detection.

The algorithm predicted each heifer's identity (Table 1) in the testing set with an average accuracy of 99.4% (i.e., ~99 out of 100 ground-truth labels for each heifer were correctly classified). The proportion of heifers cor-

Table 1. Model performance (percentage) on the image test set for each individual group-housed dairy heifer

Identification	Accuracy	Sensitivity	Specificity	PPV ¹	NPV ²
Heifer 1	99.16	98.63	99.26	96.00	99.75
Heifer 2	99.66	98.84	99.80	98.84	99.80
Heifer 3	99.24	96.05	99.62	96.81	99.53
Heifer 4	99.66	98.56	99.85	99.13	99.75
Heifer 5	99.58	98.51	99.72	97.79	99.81
Heifer 6	99.20	94.80	99.72	97.53	99.39
Heifer 7	99.45	94.95	99.86	98.43	99.54
Heifer 8	99.58	99.04	99.66	97.78	99.85

¹PPV = positive predicted value.²NPV = negative predicted value.

rectly identified by the algorithm (true positive rate), represented by the sensitivity metric, was higher than 96% except for “heifer 6” and “heifer 7.” The same 2 heifers also had a small number of ground-truth labels in the training set (“heifer 6” = 520 and “heifer 7” = 471) when compared with “heifer 1” ($n = 844$). The proportion of heifers classified as not being the target heifer (the proportion of true negatives) represented by the specificity metric was higher than 99% for all heifers. Positive and negative predicted values, which are the probabilities that a heifer was correctly identified, and a heifer predicted as 1 of the other 7 heifers was true, respectively, were higher than 96% in the testing set. Although recognizing each individual’s identity is important, the performance of such a task in scenarios where the animals are in proximity and high density (e.g., group-housed heifers) is challenging (Robie et al., 2017). However, the identification algorithm trained in this study was able to predict the individual identity of group-housed heifers with high accuracy, sensitivity, specificity, and positive and negative predicted values. The ability to predict the identity of each heifer in a group is essential to generate the behavior of interest on a specific animal, thereby moving from group behavior to individual behavior recognition (Prashanth and Sudarshan, 2020). Therefore, the development of a computer vision system such as this has the potential to be successfully applied in broader scenarios, such as animal tracking, disease detection, animal welfare, and other behavior inventories (i.e., lying, drinking, and ruminating time).

Feeding Behavior

Based on the algorithm trained to identify heifers at the feed bunk, we further predicted the NV, MVD, MIBV, and FT behaviors. The R^2 and RMSE observed for each feeding behavior predicted using 2 image every second ($n = 8,883$) are depicted in Figure 3. Low R^2 (0.39) and high RMSE (12.3, count) were observed

for NV across all heifers and time points, as shown in Figure 3A. For MVD, the R^2 was higher (0.78) than observed for NV (0.39), whereas the RMSE in minutes was 0.78 (Figure 3B). An R^2 of 0.48 and RMSE of 0.63 (min) were observed for MIBV (Figure 3C). Our algorithm predicted FT (Figure 3D) with an R^2 of 0.99 and an RMSE of 0.31 (min). The computer vision system developed in this study was capable of automatically monitoring both MVD and FT behaviors exhibited by group-housed heifers using sequential images (i.e., 1 image every second) from 4 time points. However, the algorithm did not precisely predict NV and MIBV.

Several factors may have affected the prediction of NV and MIBV, including inter-object occlusion and camera position (Chandel and Vatta, 2015; McDonagh et al., 2021). Inter-object occlusion (i.e., when part of the object of interest is occluded by another object in the scene) is 1 of the biggest challenges in computer vision, which may affect the ability to predict the output of interest (Chandel and Vatta, 2015; McDonagh et al., 2021). In this study, occlusion occurred when 1 animal mounts, rests, or stands between the camera and another animal, partially obstructing the view. In addition, the rear view of the heifers at the feed bunk (Figure 1) might have affected the ability to predict both NV and MIBV behaviors, because 1 or more heifers could cross or stop by the rear end of the heifers at the feed bunk. Changes in the light condition throughout the day can modify the appearance of an object due to shadows of different shapes and positions and shifts in the light spectrum that affect pixel intensities of each color channel (Liu et al., 1995). Lighting in the scene is considered an important factor affecting the performance of deep learning approaches, which could affect the prediction of NV and MIBV (Keller and Lohan, 2020; Hu et al., 2021). Such factors might affect the algorithm ability to detect if a specific heifer was at the feed bunk; therefore, directly influencing the NV counting, because every time the algorithm detected a heifer at the feed bunk, it considered as a new visit, and consequently reduced the MIBV. However, MVD and FT behaviors were not affected by occlusion, camera position, or lighting condition due to the short period of time the heifers were not detected by the algorithm, while they were at the feed bunk (head through the feed rail).

To tackle these problems, 1 could place the camera at the top of the barn ceiling as an alternative to the rear view used in this study, thus providing a top-down view of the animals. In addition to improving the ability to predict the feeding behavior studied here, a top-down view would also make it possible to generate other behaviors such as lying time, drinking visits, drinking time, and social interaction, among others. Such an

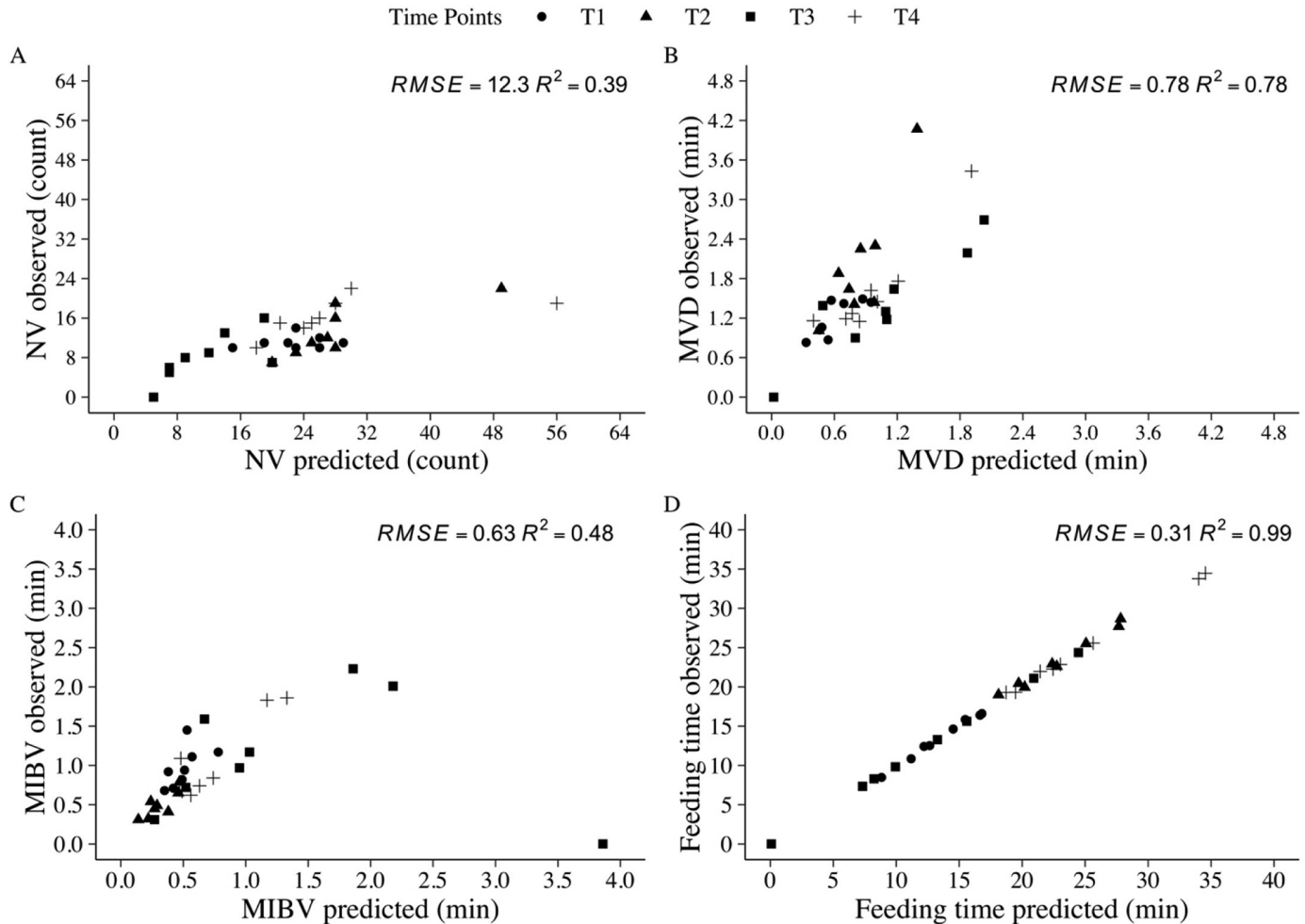


Figure 3. Observed and predicted number of visits (NV) in the feed bunk (A), mean visit duration (MVD; B), mean interval between visits (MIBV; C), and feeding time using 1 image every second for 4 time points across 8 group-housed dairy heifers. RMSE = root mean square error.

alternative is a well-known strategy to avoid occlusion in livestock systems (Psota et al., 2019); however, it is important to highlight that a camera positioned in a top-down view usually presents a reduced field of view when compared with a side-view camera angle. In this case, higher costs of implementation and analyses are expected because more cameras would need to be installed to cover the entire barn, and more data would require collection and analysis. Another alternative is to use statistical models, such as Hidden Markov Model (Ghahramani, 2001), that can model the data with sequential correlations in neighboring samples, such as in time series. This approach could detect the animal transition states between frames when an occlusion exists and predict the animal location without the need to identify it in every frame.

Instead of using the NV to infer meal frequency, DeVries et al. (2003) proposed a quantitative method to determine the total number of meals by fitting a

mixture of 2 normal distributions to the distributions of \log_{10} -transformed hit intervals. This method computes the minimum time interval between visits to determine the beginning of a new meal. The proposed approach is critical to determine the correct number of meals and related feeding behaviors, as the animal has intervals between visits within the same meal. Although the NV at the feed bunk cannot be used to describe meal behavior, its association with animal performance and efficiency (McGee et al., 2014), social interaction (Proudfoot et al., 2009), and disease (Belaid et al., 2020) has been reported in the literature. Some studies did not observe significant associations between feed efficiency and animal performance with bunk visits (Schwartzkopf-Genswein et al., 2002; Benfica et al., 2020), but reported associations between performance and FT. According to Schwartzkopf-Genswein et al. (1999), approximately 56% of bunk visits were associated with feeding activity, and the other 44% with non-feeding

activity such as scratching, licking, and rubbing. This fact partially explains the poor relationship between total bunk visits and DMI reported by other studies (Schwartzkopf-Genswein et al., 1999).

Monitoring individual group-housed heifers can provide a breakthrough in the decision-making process of dairy farm operations. In addition to its importance to nutrition management, such feeding behaviors (NV, MVD, MIBV, and FT) can be an indicator of feeding preference (Nielsen et al., 2000), respiratory diseases (Duthie et al., 2021), chemical milk composition (Gregorini et al., 2013), and early estrus (Cairo et al., 2020). Detecting individual behavior in real-time can deliver profitability, sustainability, and animal welfare benefits for dairy farm operations.

Frequency of Image Acquisition

The amount of data generated in a computer vision system is extensive, which may affect the system's efficiency, mainly due to data transfer, storage, and computational costs. Therefore, in addition to animal identification and feeding behavior prediction, we investigated the effect of different image acquisition frequencies on the ability to predict NV, MVD, MIBV, and FT behaviors. Our results showed that R^2 increased from 0.39 to 0.55, whereas the RMSE decreased from 12.30 to 2.27 (count of visits) when the image acquisition frequency for NV was decreased from 1 image every second to 1 image every 10 s (Figure 4). Surprisingly, high-frequency image acquisition did not yield the most accurate predictions of feed bunk visits (NV). As previously discussed, occlusion occurred when 1 animal mounted or stood between the camera and another animal, partially obstructing the view. In such case, the animal was not detected, but when the occlusion was eliminated, the animal was correctly identified. This process generated a new visit, as the animal was not detected in the previous frame but was in the subsequent time point (frame). With image acquisition being performed every second, the likelihood of occlusion and consequently a high number of new visits drastically increased compared with sparser image collection. After a 10-s time interval, the prediction error started increasing for NV behavior, as the true visits started to be missed due to reduced data frequency. The prediction quality for FT did not suffer from occlusion for images collected every 1 s, as the misclassification of 1 image only contributed with 1 second of error of prediction, as opposed to 1 visit in the case of NV. Using 1 image every 10 s, instead of 1 image every second, decreased both the R^2 and RMSE from 0.78 to 0.74 and from 0.78 to 0.38 (min), respectively, for

MVD (Figure 5). For MIBV, the highest R^2 (0.70) and lowest RMSE (0.22, min) were observed using 1 image every 5 s (Figure 6). Using 1 image every 10 s to predict the FT resulted in the highest R^2 (0.99) and lowest RMSE (0.43, min) compared with the prediction using 1 image every second (Figure 7). Overall, as the frequency of image acquisition decreased from 1 image every second to 1 image every 5 (MIBV) or 10 (NV, MVD, and FT) seconds, the R^2 increased, whereas the RMSE decreased for the feeding behavior studied here. Despite the possibility of using image acquisition frequencies lower than 1 image every second, our results showed that the optimal image acquisition frequency is phenotype-specific.

The frequency of image acquisition could not be decreased to lower than 1 image every 5 (MIBV) or 10 s (NV, MVD, and FT), probably, due to the natural behavior of the animals in each stage of life. Young animals tend to be more energetic, which makes them move quicker and with higher frequency, compared with the oldest animals (e.g., they quickly insert and remove their heads from the feed rail and quickly move to the next feeding space, among others). In addition, external movements, noise, or certain management practices may cause the animals to leave and then quickly return to the feed bunk. For that reason, the NV is reported to have poor relationship with some phenotypes such as feed intake and efficiency (Schwartzkopf-Genswein et al., 1999, 2002; Benfica et al., 2020). Such factors could have a direct effect on predicting the NV in the feed bunk, resulting in low prediction ability, even using 1 image every second. Likewise, the low ability to predict the NV in the feed bunk did not allow a decrease in the frequency of image acquisition. However, the feeding behaviors MVD, MIBV, and FT were not affected by increasing the frequency of image acquisition, probably because the short period that the heifers had their heads out of the feed rail did not influence the elapsed time by visit. The development of computer vision systems that optimizes model inference, training, or both is crucial to reduce the costs associated with technology implementation and data processing in farm settings.

Implications and Future Directions

Monitoring individual group-housed heifers can provide a breakthrough in the decision-making process of dairy farm operations. As opposed to several other systems that have been proposed in the literature to monitor feeding behaviors, the computer vision system developed here was able to identify each heifer at the feed bunk (head through the feed rail) with high accuracy using a single camera. The advantage of using

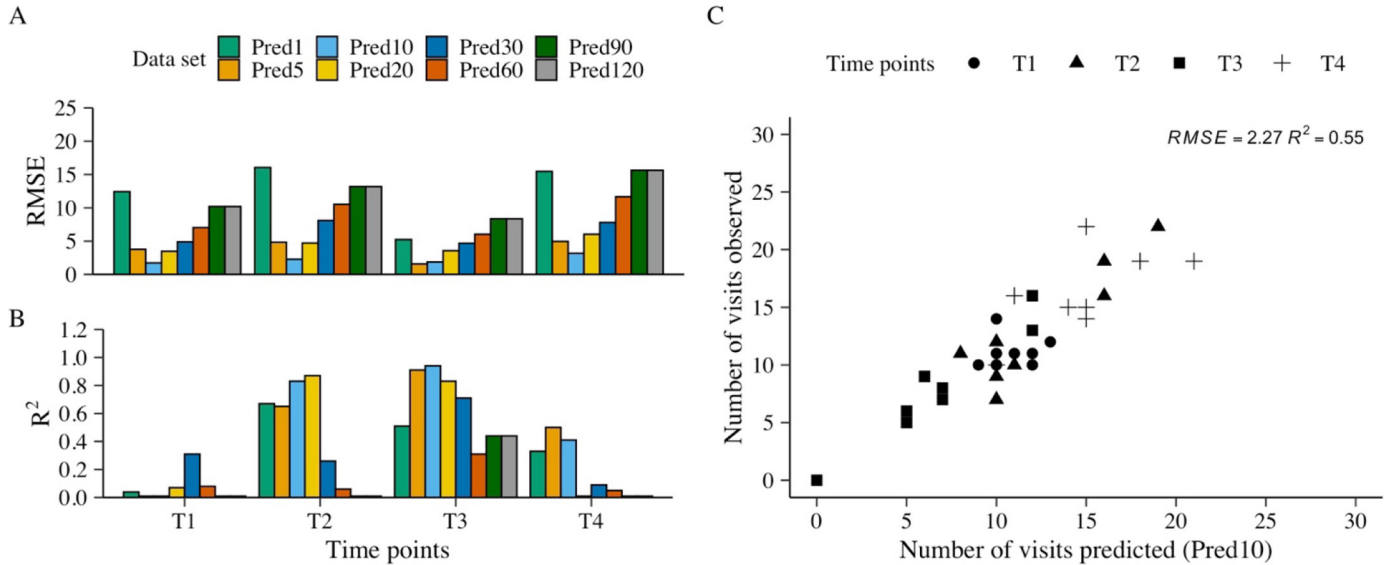


Figure 4. Root mean square error (RMSE; A) and R^2 (B) between observed and predicted total number of visits in the feed bunk using images sampled every 1, 5, 10, 20, 30, 60, 90, and 120 s (Pred1, Pred5, Pred10, Pred20, Pred30, Pred60, Pred90, and Pred120, respectively) for 4 time points (T1, T2, T3, and T3). Number of visits observed and predicted for each time point using images sampled every 10 s (C).

such computer vision system is that it does not require the use of radio frequency identification active or passive tags for monitoring feeding behavior including NV, MVD, MIBV, and FT for each heifer. Although radio frequency identification is not required for monitoring feeding behavior at the feed bunk, it could be very useful to assist image labeling (i.e., animal identification on the images), which is still a bottleneck for any super-

vised deep learning approach. Here we proposed a computer vision system based on supervised learning, but other sensors, such as radio frequency identification, could be an important component of a multi-sensor system to reliably monitor animal behavior in group-housed animals. For breeds that share the same coat color pattern, zero-shot learning and self-supervised approaches can be extremely important for monitoring

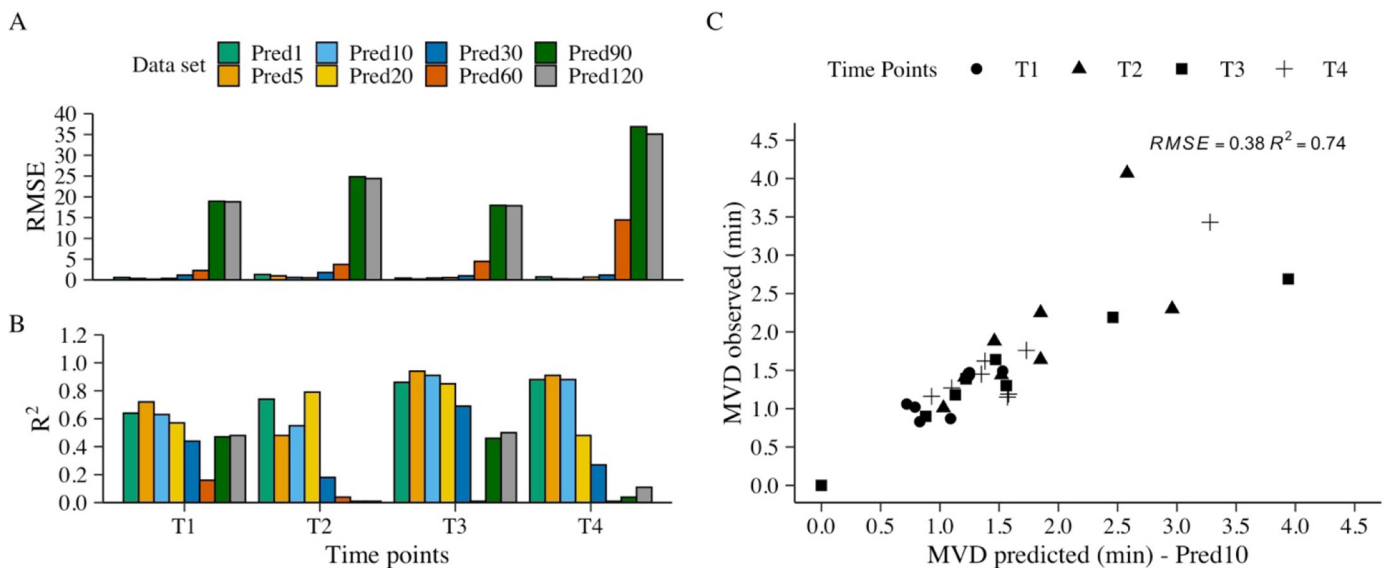


Figure 5. Root mean square error (RMSE; A) and R^2 (B) between observed and predicted mean visit duration (MVD) using images sampled every 1, 5, 10, 20, 30, 60, 90, and 120 s (Pred1, Pred5, Pred10, Pred20, Pred30, Pred60, Pred90, and Pred120, respectively) for 4 time points (T1, T2, T3, and T3). Mean visit duration observed and predicted for each time point using images sampled every 10 s (C).

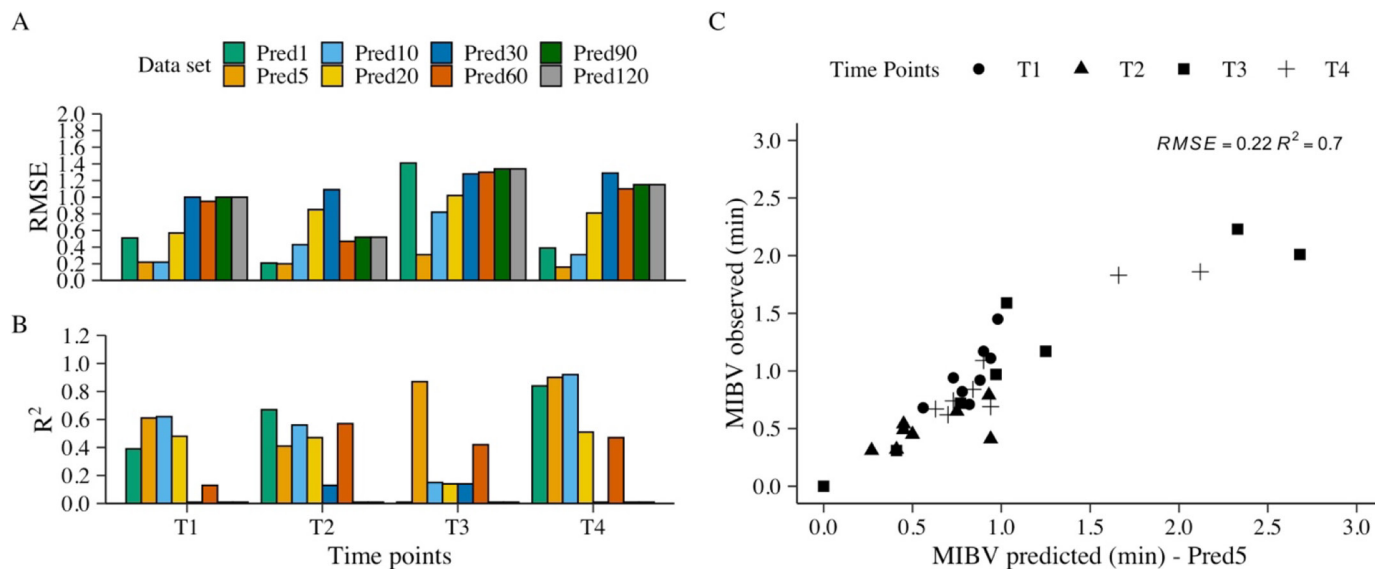


Figure 6. Root mean square error (RMSE; A) and R^2 (B) between observed and predicted mean interval between visits (MIBV) using images sampled every 1, 5, 10, 20, 30, 60, 90, and 120 s (Pred1, Pred5, Pred10, Pred20, Pred30, Pred60, Pred90, and Pred120, respectively) for 4 time points (T1, T2, T3, and T4). Mean interval between visits observed and predicted for each time point using images sampled every 5 s (C).

animal behavior (Xian et al., 2019; Dong et al., 2022). The cameras used for video-recording systems have a relatively small cost compared with wearable sensors because 1 camera can monitor a group of several animals, as demonstrated in this study, and wearable sensors can only monitor 1 animal at a time. The possibility of decreasing the frequency of image acquisition is extremely important for an efficient computer vision

system because the data size has direct effects on the costs related to computation, data transfer, and data storage. This simple but extremely important practice could result in more efficient and competitive computer vision systems that can be deployed in dairy operations, generating real-time optimal management decisions. Further investigation using computer vision system in commercial farms in addition to the research setting

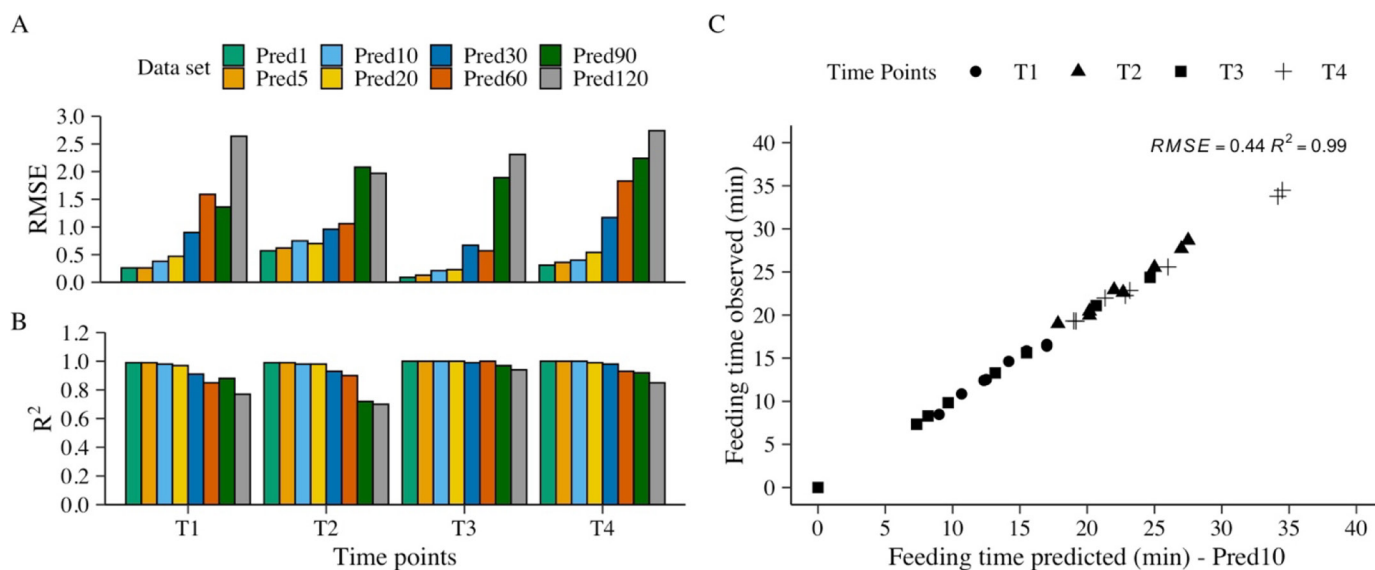


Figure 7. Root mean square error (RMSE; A) and R^2 (B) between observed and predicted feeding time using images samples every 1, 5, 10, 20, 30, 60, 90, and 120 s (Pred1, Pred5, Pred10, Pred20, Pred30, Pred60, Pred90, and Pred120, respectively) for 4 time points (T1, T2, T3, and T4). Feeding time observed and predicted for each time point by heifer (C) using images sampled every 10 s (Pred20).

used here would confirm the power of such automated system for monitoring animal behavior and beyond. The opportunity of testing the computer vision system proposed here in a broad environmental condition including lighting and longer time frame (i.e., throughout the day and night) is a need, and should be considered in future studies.

CONCLUSIONS

The computer vision system developed in this study was able to predict the identity of each group-housed heifer when they were at the feed bunk with the head through the feed rail, which is associated with feeding behavior, with high accuracy. Increasing the interval of image acquisition resulted in a better ability to predict MVD, MIBV, and FT. For NV, increasing the interval of image acquisition did not increase in the same proportion of prediction ability. The results showed that the optimal image acquisition frequencies are phenotype dependent. The importance of reducing the amount of data generated by using a lower image acquisition frequency lies in the reduction of inference time, data transfer, and data storage costs. The consideration of these factors and implementation of these computer vision aspects could improve the adoption of such technology in dairy farm operations and benefit the whole industry.

ACKNOWLEDGMENTS

The authors thank the financial support from the USDA National Institute of Food and Agriculture (Washington, DC; grant 2020-67015-30831/accession no. 1021996). The authors have not stated any conflicts of interest.

REFERENCES

- Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2015. TensorFlow. Accessed Jul. 20, 2021. <https://tensorflow.org>.
- Aikman, P. C., C. K. Reynolds, and D. E. Beever. 2008. Diet digestibility, rate of passage, and eating and rumination behavior of Jersey and Holstein cows. *J. Dairy Sci.* 91:1103–1114. <https://doi.org/10.3168/jds.2007-0724>.
- Aydin, A., O. Cangar, S. E. Ozcan, C. Bahr, and D. Berckmans. 2010. Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores. *Comput. Electron. Agric.* 73:194–199. <https://doi.org/10.1016/j.compag.2010.05.004>.
- Beauchemin, K. A., L. Eriksen, P. Nørgaard, and L. M. Rode. 2008. Short communication: Salivary secretion during meals in lactating dairy cattle. *J. Dairy Sci.* 91:2077–2081. <https://doi.org/10.3168/jds.2007-0726>.
- Belaïd, M. A., M. Rodríguez-Prado, D. V. Rodríguez-Prado, E. Chevaux, and S. Calsamiglia. 2020. Using behavior as an early predictor of sickness in veal calves. *J. Dairy Sci.* 103:1874–1883. <https://doi.org/10.3168/jds.2019-16887>.
- Benfica, L. F., L. S. Sakamoto, A. F. B. Magalhães, M. H. V. Oliveira, L. G. de Albuquerque, R. Cavalheiro, R. H. Branco, and J. N. S. G. Cyrillo. 2020. Genetic association among feeding behavior, feed efficiency, and growth traits in growing indicine cattle. *J. Anim. Sci.* 98:skaa350. <https://doi.org/10.1093/jas/skaa350>.
- Bresolin, T. 2022. TableS1. Mendeley Data, V1. <https://doi.org/10.17632/4gf2tcx8n4.1>.
- Britt, J. H., R. A. Cushman, C. D. Dechow, H. Dobson, P. Humblot, M. F. Hutjens, and G. A. Jones. 2018. Invited review: Learning from the future – A vision for dairy farms and cows in 2067. *J. Dairy Sci.* 101:3722–3741. <https://doi.org/10.3168/jds.2017-14025>.
- Brown-Brandl, T. M., G. A. Rohrer, and R. A. Eigenberg. 2013. Analysis of feeding behavior of group housed growing-finishing pigs. *Comput. Electron. Agric.* 96:246–252. <https://doi.org/10.1016/j.compag.2013.06.002>.
- Cairo, F. C., L. G. R. Pereira, M. M. Campos, T. R. Tomich, S. G. Coelho, C. F. A. Lage, A. P. Fonseca, A. M. Borges, B. R. C. Alves, and J. R. R. Dorea. 2020. Applying machine learning techniques on feeding behavior data for early estrus detection in dairy heifers. *Comput. Electron. Agric.* 179:105855–8. <https://doi.org/10.1016/j.compag.2020.105855>.
- Chandel, H., and S. Vatta. 2015. Occlusion Detection and Handling: A Review. *Int. J. Comput. Appl.* 120:33–38. <https://doi.org/10.5120/21264-3857>.
- Chollet, F., and Associates. 2015. Keras. Accessed Jul. 20, 2021. <https://keras.io>.
- Coelho Ribeiro, L. A., T. Bresolin, G. J. M. Rosa, D. Rume Casa-grande, M. A. C. Danes, and J. R. R. Dória. 2021. Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data. *J. Anim. Sci.* 99:skab206. <https://doi.org/10.1093/jas/skab206>.
- DeVries, T. J., M. A. G. Von Keyserlingk, D. M. Weary, and K. A. Beauchemin. 2003. Measuring the feeding behavior of lactating dairy cows in early to peak lactation. *J. Dairy Sci.* 86:3354–3361. [https://doi.org/10.3168/jds.S0022-0302\(03\)73938-1](https://doi.org/10.3168/jds.S0022-0302(03)73938-1).
- Dong, N., Y. Zhang, M. Ding, and G. H. Lee. 2022. Incremental-DETR: Incremental few-shot object detection via self-supervised learning. *arXiv*. <https://doi.org/10.48550/arXiv.2205.04042>.
- Duthie, C. A., J. M. Bowen, D. J. Bell, G. A. Miller, C. Mason, and M. J. Haskell. 2021. Feeding behaviour and activity as early indicators of disease in pre-weaned dairy calves. *Animal* 15:100150. <https://doi.org/10.1016/j.animal.2020.100150>.
- Ghahramani, Z. 2001. An Introduction to Hidden Markov Models and Bayesian Networks. *Int. J. Pattern Recognit. Artif. Intell.* 15:9–42. <https://doi.org/10.1142/S0218001401000836>.
- Gregorini, P., B. Dela Rue, M. Pourau, C. Glassey, and J. Jago. 2013. A note on rumination behavior of dairy cows under intensive grazing systems. *Livest. Sci.* 158:151–156. <https://doi.org/10.1016/j.livsci.2013.10.012>.
- Hu, C., B. B. Sapkota, J. A. Thomasson, and M. V. Bagavathianan. 2021. Influence of image quality and light consistency on the performance of convolutional neural networks for weed mapping. *Remote Sens.* 13:2140. <https://doi.org/10.3390/rs13112140>.
- Jiang, M., Y. Rao, J. Zhang, and Y. Shen. 2020. Automatic behavior recognition of group-housed goats using deep learning. *Comput. Electron. Agric.* 177:105706. <https://doi.org/10.1016/j.compag.2020.105706>.
- Keller, I., and K. S. Lohan. 2020. On the illumination influence for object learning on robot companions. *Front. Robot.* 6:1–10. <https://doi.org/10.3389/frobt.2019.00154>.
- Kumar, S., A. Pandey, K. S. R. Satwik, S. Kumar, S. K. Singh, A. K. Singh, and A. Mohan. 2018. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* 116:1–17. <https://doi.org/10.1016/j.measurement.2017.10.064>.

- Lin, T.-Y., M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. 2014. Microsoft COCO: Common Objects in Context. Pages 740–755 in European Conference on Computer Vision. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, ed. Lecture Notes in Computer Science, European Conference, Zurich, Switzerland.
- Liu, Y. C., W. H. Chan, and Y. Q. Chen. 1995. Automatic white balance for digital still camera. *IEEE Trans. Consum. Electron.* 41:460–466. <https://doi.org/10.1109/30.468045>.
- Mattachini, G., E. Riva, F. Perazzolo, E. Naldi, and G. Provolo. 2016. Monitoring feeding behaviour of dairy cows using accelerometers. *J. Agric. Eng.* 47:54–58. <https://doi.org/10.4081/jae.2016.498>.
- McDonagh, J., G. Tzimiropoulos, K. R. Slinger, Z. J. Huggett, P. M. Down, and M. J. Bell. 2021. Detecting dairy cow behavior using vision technology. *Agriculture* 11:675. <https://doi.org/10.3390/agriculture11070675>.
- McGee, M., J. A. Ramirez, G. E. Carstens, W. J. Price, J. B. Hall, and R. A. Hill. 2014. Relationships of feeding behaviors with efficiency in RFI-divergent Japanese Black cattle. *J. Anim. Sci.* 92:3580–3590. <https://doi.org/10.2527/jas.2013-7360>.
- Neves, P. P. 2013. Automatic registration of grazing behaviour in dairy cows using 3D activity loggers. *Appl. Anim. Behav. Sci.* 148:179–184. <https://doi.org/10.1016/j.applanim.2013.09.001>.
- Ning, H., Z. Li, C. Wang, and L. Yang. 2020. Choosing an appropriate training set size when using existing data to train neural networks for land cover segmentation. *Ann. GIS* 26:329–342. <https://doi.org/10.1080/19475683.2020.1803402>.
- Okura, F., S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, and Y. Yagi. 2019. RGB-D video-based individual identification of dairy cows using gait and texture analyses. *Comput. Electron. Agric.* 165:1–9. <https://doi.org/10.1016/j.compag.2019.104944>.
- Porto, S. M. C., C. Arcidiacono, U. Anguzza, and G. Cascone. 2015. The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system. *Biosyst. Eng.* 133:46–55. <https://doi.org/10.1016/j.biosystemseng.2015.02.012>.
- Porto, S. M. C., C. Arcidiacono, U. Anguzza, and G. A. Cascone. 2013. Computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns. *Biosyst. Eng.* 115:184–194. <https://doi.org/10.1016/j.biosystemseng.2013.03.002>.
- Prashanth, C. R., and T. S. B. Sudarshan. 2020. Deep learning methods for multi-species animal re-identification and tracking—A survey. *Comput. Sci. Rev.* 38:100289. <https://doi.org/10.1016/j.cosrev.2020.100289>.
- Proudfoot, K. L., D. M. Veira, D. M. Weary, and M. A. G. von Keyserlingk. 2009. Competition at the feed bunk changes the feeding, standing, and social behavior of transition dairy cows. *J. Dairy Sci.* 92:3116–3123. <https://doi.org/10.3168/jds.2008-1718>.
- Psota, E. T., M. Mittek, L. C. Perez, T. Schmidt, and B. Mote. 2019. Multi-pig part detection and association with a fully-convolutional network. *Sensors (Basel)* 19:852. <https://doi.org/10.3390/s19040852>.
- Rao, Y., M. Jiang, W. Wang, W. Zhang, and R. Wang. 2020. On-farm welfare monitoring system for goats based on Internet of Things and machine learning. *Int. J. Distrib. Sens. Netw.* 16:1–17. <https://doi.org/10.1177/1550147720944030>.
- Redmon, J., and A. Farhadi. 2018. YOLOv3: An Incremental Improvement. *arXiv*. <https://doi.org/10.48550/arXiv.1804.02767>.
- Robie, A. A., K. M. Seagraves, S. R. Egnor, and K. Branson. 2017. Machine vision methods for analyzing social interactions. *J. Exp. Biol.* 220:25–34. <https://doi.org/10.1242/jeb.142281>.
- Schwartzkopf-Genswein, K. S., S. Atwood, and T. A. McAllister. 2002. Relationships between bunk attendance, intake and performance of steers and heifers on varying feeding regimes. *Appl. Anim. Behav. Sci.* 76:179–188. [https://doi.org/10.1016/S0168-1591\(02\)00009-6](https://doi.org/10.1016/S0168-1591(02)00009-6).
- Schwartzkopf-Genswein, K. S., C. Huisma, and T. A. McAllister. 1999. Validation of a radio frequency identification system for monitoring the feeding patterns of feedlot cattle. *Livest. Prod. Sci.* 60:27–31. [https://doi.org/10.1016/S0301-6226\(99\)00047-0](https://doi.org/10.1016/S0301-6226(99)00047-0).
- Shaver, R. D. 1997. Nutritional risk factors in the etiology of left displaced abomasum in dairy cows: A review. *J. Dairy Sci.* 80:2449–2453. [https://doi.org/10.3168/jds.S0022-0302\(97\)76197-6](https://doi.org/10.3168/jds.S0022-0302(97)76197-6).
- Valova, I., C. Harris, T. Mai, and N. Gueorguieva. 2020. Optimization of convolutional neural networks for imbalanced set classification. *Procedia Comput. Sci.* 176:660–669. <https://doi.org/10.1016/j.procs.2020.09.038>.
- von Keyserlingk, M. A. G., and D. M. Weary. 2010. Review: Feeding behaviour of dairy cattle: Measures and applications. *Can. J. Anim. Sci.* 90:303–309. <https://doi.org/10.4141/CJAS09127>.
- Williams, M. L., W. P. James, and M. T. Rose. 2019. Variable segmentation and ensemble classifiers for predicting dairy cow behaviour. *Biosyst. Eng.* 178:156–167. <https://doi.org/10.1016/j.biosystemseng.2018.11.011>.
- Xian, Y., C. H. Lampert, B. Schiele, and Z. Akata. 2019. Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly. *IEEE Trans. Pattern Anal. Mach. Intell.* 41:2251–2265. <https://doi.org/10.1109/TPAMI.2018.2857768>.
- Yang, A., H. Huang, B. Zheng, S. Li, H. Gan, C. Chen, X. Yang, and Y. Xue. 2020. An automatic recognition framework for sow daily behaviours based on motion and image analyses. *Biosyst. Eng.* 192:56–71. <https://doi.org/10.1016/j.biosystemseng.2020.01.016>.

ORCID

- T. Bresolin  <https://orcid.org/0000-0002-3196-5150>
 R. Ferreira  <https://orcid.org/0000-0002-5141-6175>
 F. Reyes  <https://orcid.org/0000-0003-4065-1214>
 J. Van Os  <https://orcid.org/0000-0002-7107-2867>
 J. R. R. Dórea  <https://orcid.org/0000-0001-9849-7358>