Computer vision system for identification of Holstein cattle during growth and across different physiological stages

A. Negreiro[1], R. Ferreira[1], T. Bresolin[1], G. J. M. Rosa[1] and J. R. R. Dórea[1,2,*]
[1]*Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA*
[2]*Department of Biological Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*
[*]Corresponding author: João R.R. Dórea, joao.dorea@wisc.edu

Abstract

Individual animal identification is extremely important in dairy production. Accurate animal identification is not only important for farm management, but also for issues of food security and consumer trust. Computer vision has been proposed as a reliable, non-invasive system to recognize animals, but there is limited research on their capacity to detect the same individual at different stages of life. The potential for computer vision systems (CVS) to recognize an adult cow from her images as calf will allow for comprehensive animal tracking and reliable traceability systems. The objective of this study was to develop a CVS that utilizes images collected from Holstein calves in the first few weeks of life to identify the same individuals after a year of growth. To train the model, top-down view infrared images of 10 calves (1 to 4 wk of age) were collected on four separate days and segmented for the calf body, resulting in 200 images per calf. The images were used to train a deep neural network (Xception) for individual identification. The trained model was tested on 20 infrared images of each animal collected from top-down view after one year (60 wk of age). All analyses were implemented in Python using the open-source framework Tensorflow and Keras. The precision and recall of prediction for identifying individual calves were 0.78 and 0.76, respectively. These results demonstrate the potential of CVS for animal identification, even when images used for training are collected much earlier on the animals' life and during their growing phase.

Keywords: computer vision, dairy calves, identification

Introduction

Animal tracking and traceability is a matter of critical importance on dairy farms. Not only do many countries mandate individual animal identification for cattle (USDA APHIS, 2020; European Parliament and the Council of the European Union, 2000), but individual traceability throughout an animal's life can also assist in mitigating disease outbreak (USDA APHIS, 2022), and improve food security and consumer trust (Smith et al., 2005). Cattle traceability can be improved through the use of physical forms of identification such as ear-tag identification (USDA APHIS, 2020) and wearable sensors such as electronic ear tags using Radio Frequency Identification (RFID; Kang and Lee, 2013), among others. However, these identifiers can be labor-intensive on large scale operations, prone to human error, and easily lost or damaged, creating weakness in the reliability of a full-scale individual traceability system.

Computer vision has been proposed as a powerful tool to recognize and manage cattle on commercial dairy operations. Different deep learning techniques utilized in the field of computer vision allows for increased performance in tasks such as object detection, image classification, and semantic segmentation (Voulodimos et al., 2018). This creates opportunities for CVS to become an automated, non-invasive, and reliable alternative for identification of individual animals.

In recent years, individual animal identification through CVS has been an area of intensive research and innovation. Andrew et al. (2021), Bello et al. (2020), and Yukun et al. (2019) were able to accurately identify individual Holstein cows based on coat color patterns. However, these studies focused on mature cows. Recently, Ferreira et al. (2022) used depth images and 3D representation for individual animal identification in pre-weaned calves, demonstrating the potential for convolutional neural networks to identify animals over short periods of growth (6 weeks). In order to establish a comprehensive and inclusive systematic traceability system on livestock farms, it is important to investigate the ability of computer vision techniques to identify individual animals across different physiological stages as the changes in size and shape may affect their predictive performance.

Materials and methods

Data collection

For the training set, videos were recorded from 10 pre-weaned Holstein dairy calves with ages varying from one to four weeks, and body weight (BW) of 41.9 ± 4.4 kg (*average* ± *SD*), housed at the Emmons Blaine Dairy Cattle Research Center (Arlington, WI). One video was recorded for each calf once a week for four consecutive weeks using an Intel RealSense Depth Camera D435 (Intel; Santa Clara, CA) which has an RGB camera (resolution of 1920 x 1080 pixels), a depth sensor (resolution of 1280 x 720), and an infrared projector. The 40 videos were recorded from a top-down view while weighing each animal individually and contained only a single calf in each video. All videos were recorded using Intel RealSense Viewer v2.50.0 (Intel; Santa Clara, CA) installed on a laptop locally operated by a person who manually recorded each calf as it was positioned on the scale. A total of 50 images (infrared and depth) were extracted from each video at random, resulting in 200 images (with both infrared and depth frames) for each animal.

For the test set, 200 images (infrared and depth) were collected from the same 10 animals at 60 weeks of age and BW of 457.6 ± 18.7 kg (*average* ± *SD*) housed at the Marshfield Agricultural Research Station (Marshfield, WI). Heifers at this facility were housed in bedded-pack pens containing 8 heifers and a single waterer. All pens are equipped with an Intel RealSense Depth Camera D435 positioned ~4 meters above the waterer in each pen. Images were automatically collected based on motion detection during drinking time from a Jetson Nano (NVIDIA; Santa Clara, CA) connected to each camera. Images from a single day were manually cropped to include a single heifer to total 20 images for each animal.

Image preprocessing

Background removal was performed for each acquired frame in each dataset. In order to remove background pixels from the captured depth images, a network based on the Mask R-CNN (He et al., 2018) was implemented to automatically detect and retain all pixels containing an animal. For the purposes of this study, the calves' and heifers' bodies were cropped to include the region between their tails and their necks. The Mask R-CNN network for background removal was trained and leveraged from a previous study (Ferreira et al., 2022). The resulting depth mask, shown in Figure 1(b), has pixels containing the calf or heifer body appearing in white, and the background appearing in black.

For each frame, the pixels detected as containing a calf were converted to a set of points in a 3-dimensional coordinate system (a point cloud). For each pixel (i, j) containing a depth value$d$, a point (xp, yp, zp) was created with values $(xp, yp, zp) = (j, i, d)$. This resulted in a point cloud with the number of points equal to the number of pixels that were part of a calf in the original frame. Outlier points were then removed based on their Z-axis coordinates, or depth value, in order to prevent the inclusion of background pixels due to segmentation errors. A value was considered an outlier if it was more than three scaled median absolute deviations (MAD) from the median. For a random vector ***X*** with ***N*** scalar observations, the MAD is defined as follows:

$$MAD = median\,(|X_i - median(X)|)\ ,\ \text{for } i = 1, 2, \ldots, N$$

The scaled MAD is defined as $k * MAD$, where $k \approx 1.4826$ is a constant scale factor that depends on the distribution (Rousseeuw and Croux, 1993). In this case, we operated under the assumption that the Z-axis values were normally distributed.

The binary mask was then applied to the corresponding infrared image, setting every pixel not contained in the mask to zero, therefore removing the background around the animal body (Figure 1c).

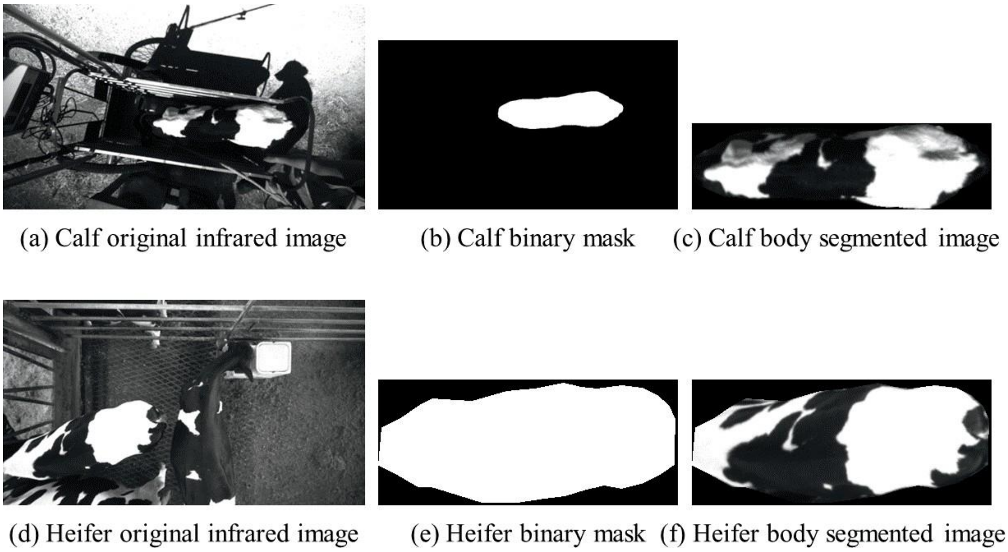

Figure 1: Example of infrared images and masks for calf (first row) and heifer (second row) body segmentation.

Table 1: Training set and test set size and animal age for experiment 1 and 2.

Experiment	Training set animal age (weeks)	Training set size (images)	Test set animal age (weeks)	Test set size
1	1-4	1,000	60	200
2	1-4	2,000	60	200

Training and test sets

Training and test sets were split by age of the animal. All images used in the training and test sets were cropped and segmented infrared images. Two different experiments were designed to evaluate the impact that training set size and age of calf on the prediction quality of the tested algorithm, shown in Table 1. In the first experiment (Experiment 1), we used 50 images per animal collected for four consecutive weeks after birth for the training set, resulting in a total of 1,000 images. In the second experiment (Experiment 2), we used 50 images per animal collected for four consecutive weeks after birth for the training set, resulting in a total of 2,000 images. Also, brightness and contrast were adjusted in training set images. The testing set for both experiments included 20 images of each animal (resulting in a total of 200 images) at 60 weeks of age housed in a different facility.

Algorithm

The deep neural network (DNN) Xception (Chollet, 2017), illustrated in Figure 2, was implemented in Python using the open-source frameworks TensorFlow (Abadi et al., 2015) and Keras (Chollet et al., 2015). The last Fully-Connected (FC) layer of the original architecture was removed, and all the other layers were initialized with weights trained using the open image dataset ImageNet (Deng et al., 2009). This strategy, defined by Weiss et al. (2016) as Transfer Learning, accelerates the training process by initializing the network weights with values optimized for a large generic image dataset instead of random values.

The Xception-based DNN was extended with a global average pooling layer as described by Lin et al. (2014), followed by a FC layer of size 1024 and the ReLU activation function, then a final FC layer of size *n* and the softmax activation function.

The training process was split into two consecutive stages: feature extraction and fine-tuning. In the feature extraction stage, the DNN was trained for 200 epochs keeping the weights of all but the last two FC layers frozen. This allowed features previously learned through Transfer Learning to be used and retained. In the fine-tuning stage, weights from earlier layers were unfrozen, and the network was trained for 400 epochs with a smaller learning rate, allowing it to further learn features that are more specific to our context.

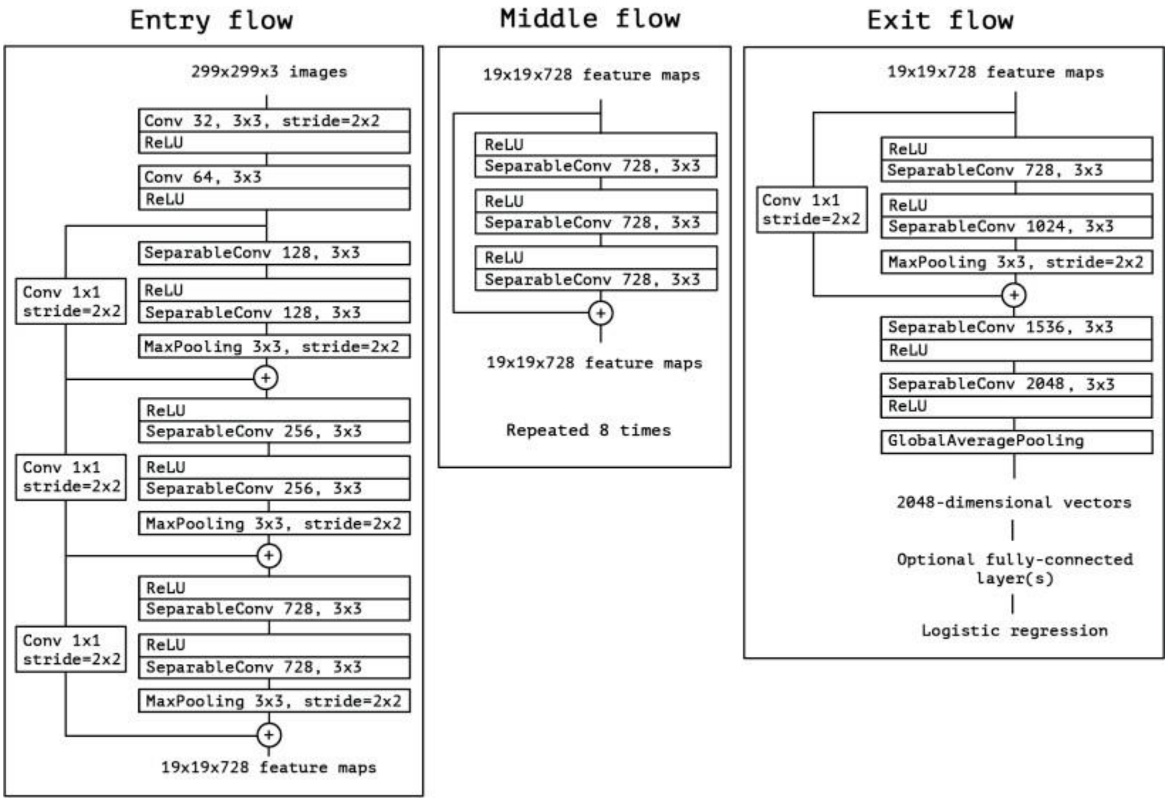

Figure 2: Representation of the Xception architecture. Adapted from "Xception: Deep learning with depthwise separable convolutions." by Chollet, 2017.

Evaluation metrics

To evaluate the prediction quality of the algorithms, the accuracy, precision, recall, and F_1 score were calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

where: TP = True positives, TN = True negatives, FP = False positives, FN = False negatives.

Results and discussion

In Experiment 1, the accuracy, precision, recall, and F_1 score of prediction for identifying individual calves were 0.36, 0.37, 0.33, and 0.35, respectively. In Experiment 2, the same metrics were 0.76, 0.78, 0.76, and 0.77, respectively.

Table 2: Accuracy, precision, recall and F1-score for experiments 1 and 2

Experiment	Accuracy	Precision	Recall	F_1 score
1	0.36	0.37	0.33	0.35
2	0.76	0.78	0.76	0.77

The large increase in prediction performance from Experiment 1 to Experiment 2 demonstrates that larger training sets are necessary for individual animal identification during periods of growth or across different physiological stages in dairy cattle. Similarly, Ferreira et al. (2022) illustrated the importance of training set size. For pre-weaned calves, performance of multiple convolutional neural network architectures improved as the number of training images per animal increased, up until 100 images per animal (Ferreira et al., 2022). In contrast, Nye et al. (2020) found that a sample size of only 50 images were required to train their semi-supervised machine learning approach for dairy cattle identification. Importantly, Nye et al. (2020) used adult cattle and side-view images, while our experiments were designed to identify animals after physiological changes from top-down view.

The results from Experiment 2 show that using infrared images from calves in their first month of life can be used to identify them after a year and in a different environment. This is an important step in individual animal traceability on dairy operations with multiple locations. Although individual animal identification based on coat color pattern has been proposed by Andrew et al. (2021), Bello et al. (2020), and Yukun et al. (2019), these studies focused on mature cattle and did not include identification across different physiological stages. In order to establish traceability of cattle throughout their lives, development of systems to reliably identify the same animal regardless of physiological change is crucial. To the best of our knowledge, this is the first work to evaluate the ability of convolutional neural networks to identify animals over long periods of growth and across different physiological stages.

However, there are limitations to deep learning methods that rely on coat color pattern, as their performances are greatly decreased when applied to recognize individuals in dairy and beef cattle breeds with homogenous appearances such as Jersey, Brown Swiss, and Angus. These methods are also limited due to lighting, significant occlusion, and scenarios where animals can be covered in mud or dirt. An approach to address these limitations was proposed by Ferreira et al. (2022), where depth images were used to identify individual pre-weaned Holstein calves within 6 consecutive weeks (F_1 scores > 0.80). Although this identification strategy performed well with images collected from calves at a young age, it is unknown how well it would perform across long periods of growth. The next step in this study is to combine and compare 2D and 3D methods of animal identification during growth and across different physiological stages to create a robust system for individual animal traceability.

Conclusions

Computer vision systems are a promising solution to individual animal traceability through periods of growth and physiological changes in dairy cows. An Xception-based deep neural network performs with high precision and recall when identifying individuals with unique coat patterns after one year of growth.

Acknowledgments

This research was performed using the computational resources and assistance of the University of Wisconsin-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by University of Wisconsin-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. The authors thank the financial support from the USDA National Institute of Food and Agriculture (Washington, DC; grant 2020-67015-30831/accession no. 1021996) and NVIDIA "Academic Hardware Grant Program".

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., and Devin, M. (2015) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *ArXiv* 1603.04467.

Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A.W., and Burghardt, T. (2021) Visual identification of individual holstein-friesian cattle via deep metric learning. *Computers and Electronics in Agriculture* 185, 106133.

APHIS, U. (2020) *Cattle Identification*. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/nvap/NVAP-Reference-Guide/Animal-Identification/Cattle-Identification

APHIS, U. (2022). *Animal Disease Traceability*. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/traceability

Bello, R.-W., Talib, A.Z., Mohamed, A.S.A., Olubummo, D.A., and Otobo, F.N. (2020) Image-based individual cow recognition using body patterns. *Image* 11(3), 92-98.

Chollet, F. (2015) *Keras, Github*. https://github.com/fchollet/keras

Chollet, F. (2017) Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1251-1258.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fe

i, L. (2009) Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248-255.

Ferreira, R.E., Bresolin, T., Rosa, G.J., and Dórea, J.R. (2022) Using dorsal surface for individual identification of dairy calves through 3d deep learning algorithms. *Computers and Electronics in Agriculture* 201, 107272.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017) Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision* 2961-2969.

Kang, Y.-S., and Lee, Y.-H. (2013) Development of generic rfid traceability services. *Computers in Industry* 64(5), 609-623.

Lin, M., Chen, Q., and Yan, S. (2013) Network in network. *ArXiv* 1312.4400.

Nye, J., Zingaretti, L.M., and Pérez-Enciso, M. (2020) Estimating conformational traits in dairy cattle with deepaps: A two-step deep learning automated phenotyping and segmentation approach. *Frontiers in Genetics* 11, 513.

Rousseeuw, P.J., and Croux, C. (1993) Alternatives to the median absolute deviation. *Journal of the American Statistical Association* 88(424), 1273-1283.

Smith, G.C., Tatum, J.D., Belk, K.E., Scanga, J.A., Grandin, T., and Sofos, J.N. (2005) Traceability from a us perspective. *Meat Science* 71(1), 174-193.

Union, E.P.a.t.C.o.t.E. (2000). *Establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing council regulation*. https://eur- lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000R1760

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018) Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience* 2018.

Weiss, K., Khoshgoftaar, T.M., and Wang, D. (2016) A survey of transfer learning. *Journal of Big Data* 3(1), 1-40.

Yukun, S., Pengju, H., Yujie, W., Ziqi, C., Yang, L., Baisheng, D., Runze, L., and Yonggen, Z. (2019) Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. *Journal of Dairy Science* 102(11), 10140- 10151.