Applied Animal Behaviour Science 292 (2025) 106746

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

Exploring pose estimation as a tool for the assessment of brush use patterns in dairy cows

Niclas Högberg [a,*], David Berthet [b], Moudud Alam [c], Per Peetz Nielsen [d], Lena-Mari Tamminen [a], Nils Fall [a], Adrien Kroese [a]

[a] Department of Clinical Sciences. Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural sciences, Ulls väg 26, Uppsala 756 51, Sweden
[b] Sony Nordic (Sweden), Mobilvägen 4, Lund 221 88, Sweden
[c] School of Information and Engineering, Dalarna University, Högskolegatan 2, Falun 791 88, Sweden
[d] RISE Research Institute of Sweden, Division of Bioeconomy and Health, Department of Agriculture and Environmental Engineering RISE Ideon, Lund 223 70, Sweden

ARTICLE INFO

Keywords:
Behaviour
Monitoring
Pose Estimation
Welfare indicator

ABSTRACT

Access to mechanical brushes enables grooming behaviour in dairy cows and has shown benefits for cow welfare, including improved cleanliness, comfort, stress reduction. Brush-use may also promote a positive emotional state. Reduced brush use has been associated with health issues, suggesting its potential for automated health monitoring. This study aimed at evaluating whether data generated by pose estimation could be used to assess brush use patterns in loose-housed dairy cows. It presents an approach for automatically identifying the body segment being brushed as an application of pose estimation. Data collection was carried out at the Swedish Livestock Research Centre in a loose housing system equipped with an automatic milking system and two mechanical rotating brushes. Recordings spanned 25:30 h and used three cameras, at different positions, monitoring a single mechanical brush placed in a passageway between cubicle rows. One human observer with access to recordings from all three synchronized cameras annotated the data-set on a second-by-second basis. The observer recorded: (1) the number of cows using the brush; (2) the anatomical segment being brushed; and (3) whether brushing resumed after a pause. The same video recordings were processed with object detection and pose estimation, which predicted the location of bounding boxes for cows and for the brush as well as corresponding keypoints. Using the brush and cow keypoint locations, we attempted to detect brushing by anatomical region. In a first stage, machine-learning models were trained to predict brushing state (independent of location) using keypoint distance to the brush, achieving an accuracy of 86.3 %. To mitigate the risk of error propagation, we relied on human annotations to segment the video to confirmed brushing bouts for analysis in the second stage. To identify the anatomical location of brushing, two methods were evaluated: (1) simply assigning the brushing location to the closest keypoint, achieving 73 % average accuracy across classes, and (2) projecting brush and anatomical keypoints onto a spline modelling the cow's backline, resulting in 87 % accuracy. Misclassifications were predominantly limited to adjacent body segments. Given that intra-observer reliability was 90 %, the spline-based method was deemed sufficiently reliable for research applications to accurately monitor the specific body segments being brushed.

1. Introduction

The access to mechanical brushes facilitates natural grooming behaviour and has shown promising results in improving various aspects of cow welfare, including cleanliness, stress reduction (DeVries et al., 2007), comfort (EFSA Panel on Animal Health and Animal Welfare et al., 2023), and has been suggested to promote a positive emotional state (Keeling et al., 2021). Cows demonstrate high motivation to use mechanical brushes, and it has been reported that within one week after the installation of a brush, a majority (93 %) of cows use it (DeVries et al., 2007). Moreover, it has been observed that cows are willing to push as much weight to access a brush as they would to access fresh feed (McConnachie et al., 2018).

Animals allocate time and effort to a range of luxury (e.g., playing) activities. A luxury activity is characterized by low resilience and will be reduced when time or energy are limited (McFarland, 1999). As

* Correspondence to: SLU, Department of Clinical Sciences, Box 7054, Uppsala 750 07, Sweden.
E-mail address: Niclas.hogberg@slu.se (N. Högberg).

https://doi.org/10.1016/j.applanim.2025.106746
Received 3 December 2024; Received in revised form 16 June 2025; Accepted 9 July 2025
Available online 22 July 2025

grooming is considered an activity with a low resilience, the use of mechanical brushes has the potential to be implemented as a method for automated health monitoring (Littin et al., 2008; Weary et al., 2009). Previous research is still limited but has found reduced brush use in cattle with metritis (Mandel et al., 2017) and lameness (Burton and Blackie, 2024; Weigele et al., 2018).

The access to brushes may also promote positive emotional states. Cows show relaxed behaviours and postures in connection to brush use (De Oliveira and Keeling, 2018). Engagement in brush use tend to increase in non-stressful times and decreases during stress (Lecorps and Féron, 2015; McConnachie et al., 2018). This highlights the crucial role of the brush in enhancing the welfare and comfort of cows, emphasizing its value in reducing stress and promoting positive emotional states.

Grooming of different body regions, in the form of human stroking, has been linked to different physiological and behavioural responses. Stroking of the withers region resulted in a lower heart rate and longer neck stretching and ear hanging, behavioural indicators of relaxation, compared with other body regions (Schmied et al., 2008). A recent study examining how stress (head restraint) influences brush use found that cows in the control group were more likely to brush their withers region. This indicates that the body region cows choose to brush might vary based on their emotional state, suggesting that brushing of the withers region could be linked to positive emotional states (Skånberg et al., 2017). Conventional methods for monitoring brush interaction, such as direct or video observation, are labour-intensive, limited in scale, and not feasible for implementation in commercial settings. To enable scalable and real-time monitoring, automated approaches are necessary. The development of automated methods to enable on-farm data collection to assess animal welfare was recently highlighted in a scientific opinion from the EFSA Panel on Animal Health and Animal Welfare (2023).

Previous experimental approaches to automatically record brush use have focused on the use of radio frequency identification (RFID) (Toaff-Rosenstein et al., 2017) and real-time location systems (RTLS) (Meunier et al., 2018). For RFID, mean sensitivity and specificity were 0.88 and 0.91, respectively, but varied between individuals. Accuracy reached 90.5 %. In comparison, the RTLS system achieved accuracy, negative predictive value, and true negative rate values above 75 % for detecting brush use. Researchers applying the tool to derive conclusions on animal behaviour need to decide whether the error rate is within acceptable bounds. A more recent study highlighted the possibilities of machine learning approaches on fused sensor data channels, combining automated brush rotation logging with fiducial markers and computer vision, or RFID, for identification (Sadrzadeh et al., 2024). However, for bout detection, logistic regression surpassed machine learning approaches with a reported precision of 0.84, and a recall and F1-score of 0.90 and 0.87, respectively. In the same study, random forest was most effective at identifying the true user among cows detected in close temporal proximity, while a multilayer perceptron performed best at excluding incorrect users, based on both RFID and fiducial marker data.

Computer vision is a field of artificial intelligence that enables machines to interpret the visual world by analysing for example images or video (Fernandes et al., 2020; Szeliski, 2011). It has the advantages of being an objective, non-invasive and continuous method and it has been applied for recognition of different livestock behaviours (Chen et al., 2021). Pose estimation is a computer vision technique that involves predicting and tracking body posture by localizing anatomical key points. 2D pose estimation predicts the location on a 2D grid (X, Y), whereas 3D pose estimation infers the spatial position by adding an extra Z-axis to the predicted key points (Ben Gamra and Akhloufi, 2021). Single-camera (2D) setups are cost-effective and easy to install but have a limited field of view, face occlusion issues, and lack redundancy, potentially missing key behaviours. Multi-camera (3D) setups offer better coverage, improved accuracy, and enhanced behavioural analysis by handling occlusions effectively. However, they require higher costs, more complex installation, and greater data management. Several studies have been published for animal 2D pose estimation (Mathis et al., 2018; Pereira et al., 2019; Russello et al., 2022). Recently a 3D pose estimation approach was successfully used to detect posture transition in freestall-housed dairy cows (Kroese et al., 2024).

Pose estimation of cows has successfully been applied, including in situations of high occlusion (Gong et al., 2022), notably by retraining existing models on annotated cow images (Li et al., 2019). These models provide important information on the spatial location of anatomical structures, and generate continuous information at a scale and consistency which would be unrealistic with traditional human annotations. This information is useful but offers limited insights in itself; the movements of the keypoints need to be interpreted from a behavioural perspective in order to draw meaningful conclusions. This study represents an attempt to utilize the output from pose estimation into a framework for brushing behaviour monitoring.

Our objective was to explore the use of 2D pose estimation in assessing mechanical brush use patterns in dairy cows. Specifically, we investigated whether proximity, i.e. the distance between the brush and key anatomical landmarks, could reliably identify which body segment was being brushed at a specific time. To determine the most effective configuration, we tested three camera angles and compared their performances separately. This study details the development of the method and quantifies its agreement with manual annotations. It provides descriptive results in the form of brush bout characteristics and brushing locations to illustrate the output generated.

2. Materials and methods

2.1. Ethical considerations

The study was approved by the Swedish Board of Agriculture's Uppsala Ethics Committee on Animal Research and performed according to the Swedish legislation on animal experiments (diary number 5.8.18–03052/2024).

2.2. Animals and housing

The data used in this study was recorded at the Swedish Livestock Research Centre, Uppsala, Sweden. The lactating cows were loose housed in a voluntary milking group (VMS™ 300 DeLaval International AB, Tumba, Sweden). They were provided a roughage mix consisting of grass-clover and corn silage, individual rations of concentrate in transponder-activated feeders and water *ad libitum.* The herd consisted of Swedish Red (SR) (60 %) and Swedish Holestein (SH) (40 %). The data was recorded between March 19, 2024 and March 20, 2024. The group had access to two mechanical rotating brushes (DeLaval swinging cow brush SCB, DeLaval, Tumba, Sweden). A schematic lay-out of the group can be seen in Fig. 1. At the time of data collection, the group consisted of 50 cows (SH=30, SR=20).

2.3. Data collection

Continuous video recordings were obtained from three synchronised RGB cameras (G3 Bullet, Ubiquiti Inc., New York, United States). The cameras were placed at different locations in connection to one mechanical brush in the group (Fig. 1): (1) Top-down: 4.5 m; (2) Side view 3 m; (3) High angle 4 m. The top-down view was chosen so that all cows and the brush would be visible at all times, without occlusion. The side view was assumed to produce the most interpretable data, as when a cow is brushing, its back would be parallel to the X-axis of the frame and the brush would be located along that axis which would enable straightforward calculation of the brush positions relative to the cow. The high-angle camera was placed for convenience. The video recordings were timestamped and stored in 1-hour sections.

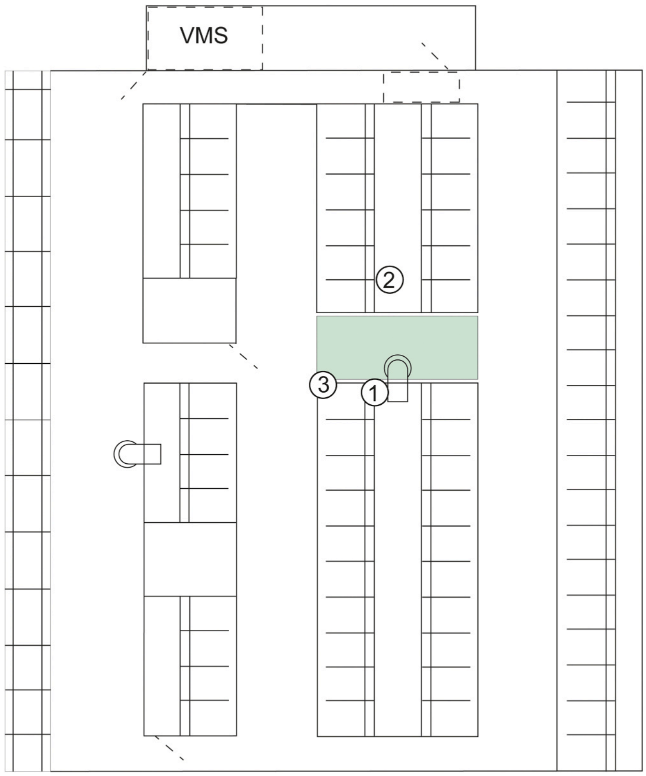

Fig. 1. Top-view schematic overview of experimental group. Mechanical brushes, cubicles, slatted aisles, the VMS and feed bins are given. The area of interest was video recorded is highlighted by the green box. Camera placements and height are indicated by: (1) Top-down 4.5 m; (2) Side view 3 m; (3) High angle 4 m.

2.4. Human observation and annotation

Human observations served as the ground truth for our analysis. One trained observer recorded brushing behaviour continuously for a total of 25:30 h from the top-down camera. To enable assessment of intra-observer reliability the observer scored 4 h of video two times with 23 days in between. The video from all three cameras was available for the observer. The observer recorded in Excel (Microsoft Corporation, version 16.0) the start and finish timestamps of physical contact of any duration with the brush when it was spinning, the part of the body being brushed, whether the brush was used by the same cow as in the previous interaction (based on visual identification of the individual), and the number of cows brushing. To record which part of the body the cow was brushing, seven body segments were defined based on the pre-defined keypoints generated by the pose estimation model (Fig. 2). These included the muzzle, poll, neck ((7th cervical vertebrae (C7)), highest point at the withers, the 13th thoracic vertebrae (T13), sacrum taken immediately behind the uppermost part of the ilium and base of the tail.

Brushing location on the body was determined in the following ways:

- If the brush was located somewhere between the head and the tail, the observer reported the keypoints enclosing the brush. For example, if the brush was detected closest to the withers, the observation for brushing location would be "between neck and T13", as is exemplified on Fig. 2.
- If the brush was beyond the tail or muzzle, then the annotation would be "beyond the sacrum" or "before the muzzle".

Each time a different segment was brushed a new timestamp was recorded. If the brush was moving between segments at a rate higher than one segment per second, the observer recorded it as *no specific segment*, thereby indicating that the cow was using the brush, albeit with no noted position. For the cases where more than one cow used the brush simultaneously, the observer assigned a specific individual ID (i.e. 1, 2, 3, etc.) to the cows to enable differentiation between individuals. It should be noted that the observer did not record the individual identification of the cows observed.

Nearly four consecutive hours of recording were re-assessed to measure intra-observer agreement. The annotations resulted in intervals of varying lengths where the brush-use status was constant. To compare them, we expanded the entire observation period into one-second increments, creating a common timeline for both sets of annotations. The agreement was checked for each second of recording, noting first whether the observer agreed on brush-use and then within brush-use bouts if they agreed on brush-use anatomical location.

Prior to computing agreement on brush-use segment, time intervals labelled as *no specific segment* were excluded from the analysis to focus on periods of brushing on a clear segment. We employed a penalty-based approach to quantify agreement, accounting for both temporal overlap and the anatomical distance between regions in cases of disagreement; the further along the cow's length two annotations, the higher the penalty. Penalties increased linearly with the order of the keypoints along the cow's axis. Penalty weighted agreement was calculated as:

$$1 - \frac{\sum_{t=00:00:01}^{03:57:29} P_t}{t_{max}}$$

Where t represents the timestamp, t_{max} is the total duration re-annotated by the same observer, and P is the penalty. Penalty increased linearly between successive keypoints. The penalty was 0 in case of agreement on annotated segment for a timestamp and took a value of 1 if two annotations for the same timestamp were labelled as adjacent keypoints, and for the most extreme case 6 if the most distant possible keypoints (i.e. muzzle and tail) were selected. This result provides the percentage of timestamps where both series of annotations matched, with greater mismatches contributing multiple counts for the same timestamp.

Fig. 2. Selected frame depicting two cows, with one using the brush (confirmed by human annotations). Keypoint proximity to the brush assigned the correct object (green) to the cow using the brush. Object detection and pose estimation predicted keypoints and bounding boxes that are overlaid. Keypoint legend from bottom-left: muzzle, poll, neck, withers, T13, sacrum, tail base.

2.5. Pose estimation

The recordings from all three cameras were processed separately with object detection and pose estimation. Fig. 4 provides an overview of the entire process of both annotations, detection of brushing, and brushing location using either method.

Object detection for cows and brushes was done using YOLOX (Ge et al., 2021). Regions of interest (ROI) were manually defined on the cameras' fields of view and corresponded to the corridor where the brush was located (green in Fig. 1). This decreased the computation requirements by constraining the object detection to areas that are relevant to the brush study (excluding cubicles and alleyways which are not relevant to this study). YOLOX is a deep learning-based object detection model that efficiently identifies and locates animals (or other objects) within images or video frames by predicting bounding boxes and class labels directly in a single pass, making it useful for real-time monitoring and analysis in animal research. Cow poses were estimated using HRNet (Wang et al., 2019). HRNet is a deep learning model for pose estimation — that is, the detection of anatomical features of animals. Unlike YOLOX, which progressively compresses the image during convolutional steps to extract high-level features, HRNet maintains detailed spatial information throughout the network by preserving high-resolution representations, allowing it to precisely detect fine-grained details such as joints or other anatomical landmarks, which is especially valuable for analyzing fine movements and posture changes. Both models were retrained on proprietary images showing cows in a diversity of environments, on which keypoints corresponding to anatomical landmarks were annotated. The keypoints used in this study corresponds to the following anatomical landmarks: muzzle, poll, neck (C7) highest point at the withers, T13, sacrum taken immediately behind the uppermost part of the ilium and base of the tail. Because of few interactions with the brush at the muzzle, occurrences at the muzzle and poll keypoints were merged as "head". The centre point of the brush bounding box provided by the object detector was used to represent the brush in its 2D context. An example of pose estimation and object detection can be seen on Fig. 2.

2.6. Data post-processing and brush use detection

The eventual purpose of pose estimation was to determine which body parts cow were brushing. This includes three subsequent steps which are detailed in this sub-section: (i) determining whether a cow is brushing, (ii) determining which cow is brushing when there are several on the scene and (iii) identifying which anatomical segment is being brush

Cow poses were estimated on all frames from the continuous 25:30 h of recordings (1777 949 frames at a down-sampled rate of 2fps), for all objects in the ROI, while annotations were provided in the form of time intervals for each cow and at each segment being brushed. Detections and annotations were merged based on timestamp-interval correspondence. This led to some video sequences only having one cow annotated (because a single one was brushing) but several objects detected (since there could be several cows in the scene). Therefore, it was necessary to assign the right detection to the individual using the brush. This was done by calculating the distance between each object's keypoints and the brush (precisely, the centre of the brush bounding box). We assigned the object with the smallest overall keypoint-to-brush distance to the animal brushing.

To validate this method, keypoints and bounding boxes were overlaid onto 100 randomly selected frames showing several poses and having one animal annotated as brushing. On each frame, the bounding box of the object which had the closest keypoint to the brush was highlighted. It was visually checked whether this object did truly correspond to the animal using the brush. Other objects present were discarded from the analysis. An example is shown on Fig. 2.

Prior to analysing brush-use patterns, it was necessary to detect interactions with the brush and define bouts. Successive interactions with the brush were grouped into the same bout if they were separated by less than 25 s. After grouping, some bouts were removed if they lasted less than 5 s or if no clear body region was identified as brushing throughout the bout. This was indicated in the annotations as *no specific segment*. The threshold of 25 s was determined from the distribution of brush-use interruptions. Fig. 3 A shows the distribution of the durations separating two successive brush interactions by the same animal. A sharp elbow in the distribution, followed by stabilization after 25 s led us to

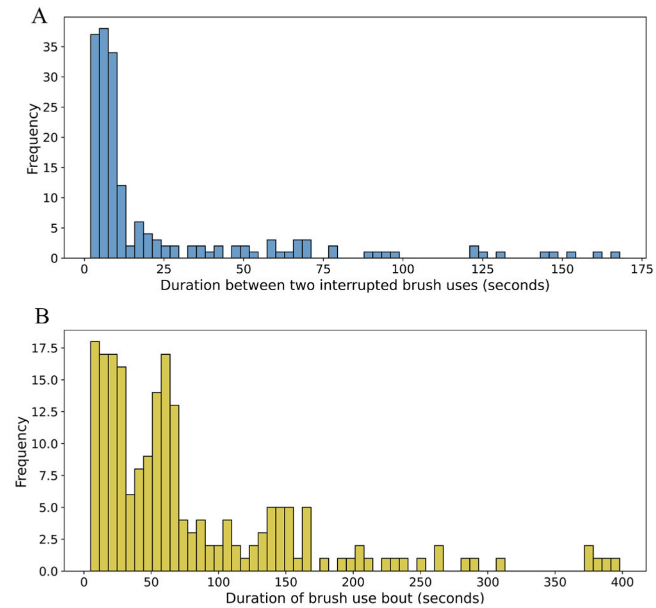

Fig. 3. Histograms of brushing bout duration and interruptions. Upper 5th percentile removed for readability.

consider two successive interactions separated by a shorter interval as likely corresponding to the same bout.

The annotations regarding brush use status were merged with the poses by frame, to create labels indicating at each frame whether a cow was brushing or not. All sequences with more or less than one cow (based on object detection output) were removed for this stage. Brush use has a temporal dimension, that is, when a cow begins and finishes using the brush. The aim of this section was to accurately detect the start and end of brushing bouts, encompassing a temporal dimension rather than frame by frame, while maintaining a high level of interpretability by using brush to skeleton distance.

The Euclidian distance in pixels between each keypoint and the brush was calculated and averaged into rolling windows of 5 s. The brushing state was set to *True* if any frames within the window were annotated as brushing and *False* otherwise. We attempted to predict the brushing state for each 5 s window using either a random forest or a recurrent neural network.

The data was split into the first 70 % of frames for training and the last 30 % for testing. Such splitting was done to maintain time-continuity and avoid data leakage in the test set. The testing set contained 40 separate brushing occurrences with durations between 3 s and 20 min (corresponding to 5th and 95 % percentile of the duration distribution). It was thus considered diverse enough for validation.

The random forest trained with 37 trees resulted the highest accuracy in predicting brushing state. The random forest was fit using the default parameters of the function RandomForestClassifier in the scikit-learn package (Pedregosa et al., 2011). The neural network was implemented in the keras package (Chollet, 2015) and was parametrized with a batch size of 16, learning rate of 0.001 and strides of two frames, and contained the following layers:

Recurrent layer with long short-term memory (LSTM) with a length of 32 windows and tanh activation,

Dropout of 0.2,

Dense layer with 16 *units* and relu activation,

Dense layer with one *unit* and sigmoid activation.

We compared the brushing status predicted by the models at each window against the annotations and computed overall accuracy. 37.5 % of frames showed brushing, thus creating a fairly balanced dataset, justifying the use of overall accuracy.

In the next section, we describe how the animal was segmented based

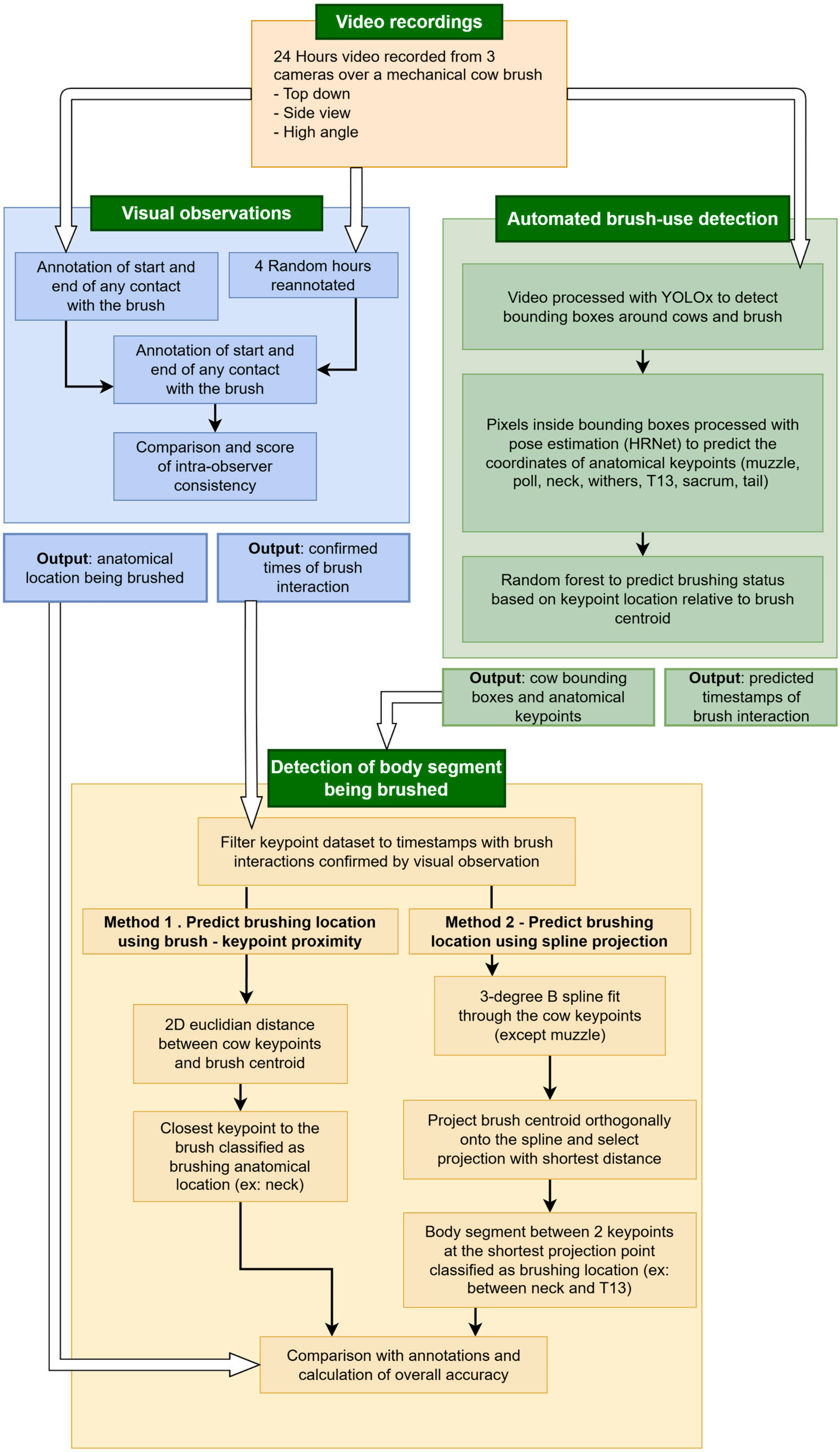

Fig. 4. Flowchart outlining the applied method, including annotation, brushing detection, and location identification.

on the keypoint location and how the brush position on the body was determined. To ensure that this process was done on data accurately reflecting the ground truth (brushing) status, it was decided to segment the bouts to known time windows where an interaction with the brush was occurring, based on the human observations.

2.7. Determining brushing segment using keypoints

Detection of brush bouts was tested algorithmically using the method described in the previous section. To limit the risk of error propagation across steps, confirmed brush bouts were defined based on timestamps annotated by the observer, not those predicted. A total of 171 events were used.

2.7.1. Using the closest keypoint to the brush

To identify where on the body the brush is being used, we first employed a coordinate-based approached, leveraging the relative distance of the brush to the selected keypoints. The Euclidian distance in pixels was first calculated between the brush and each keypoint, both represented by x and y coordinates on the 1980*1080 frame. The classification method that was applied for the human observer (see 2.4. Human observation and annotation) was modified for automated detections. The classification was based on the keypoint with the shortest distance to the brush location, which was considered the brushing location.

2.7.2. Using a projection on a spline fitted to the keypoints

The second method involved fitting a degree-3 B-spline with no smoothing, using the cow keypoints as knots to approximate the animal's backline. The muzzle was ignored when fitting the spline because of its spatial proximity to the head and the low occurrences of muzzle-brush interactions. The spline was standardized from 0 to 1 between the head and the tail, allowing us to compute relative positions along this curve. To determine where the brush was positioned relative to the animal, we projected both the cow keypoints and brush on the spline. The projection on a spline enabled to determine between which two keypoints the brush was located, thus mirroring the way the annotations were done. If the brush was ahead of the head of beyond the tail, the brushing location was considered respectively "head" or "tail". Within the segments between two keypoints on the spline containing the projection of the brush, the distance was computed between the brush projection and the two surrounding keypoints. Custom cut-offs for determining which of the two keypoints was being brushed were determined by a decision tree classifier of depth 1 for each segment.

2.7.3. Assessment of agreement

The effectiveness of this classification approach was evaluated by comparing the predicted locations against human annotations using metrics for precision, recall, and macro-average accuracy (average accuracy over each class independent of their sizes). Since this classification is performed using analytical rules rather than supervised learning – relying on the high interpretability of pose estimation – and is not trained on annotations, accuracy was computed on the entire dataset rather than a test subset. For comparability between annotations and detections, annotations were relabelled to reflect the keypoint in the annotated segment. For example, the annotation "between neck and T13" was relabelled as "withers".

3. Results

3.1. Agreement on brush use

Intra-observer agreement on brush use was 91 % (in percentage of annotated frames). Agreement between the observer and the models on brush use was 86.3 % for the random forest and 83.9 % for the neural network after stabilisation of the loss at 200 epochs. F1 score was 0.81. When two cows were present in the region of interest but only one of them was using the brush, the correct cow was identified as being the brush user in 98 % of randomly sampled frames. We conclude that assigning the brushing object to the animal closest to the brush was sufficiently robust to proceed with the analysis.

3.2. Agreement on body segment being brushed

For the human annotation of segments, both raw agreement and penalty-weighted agreement for keypoint location were 90 %. All disagreements were on adjacent keypoints and the most common disagreements were between neck and withers, and between neck and head. Sources of disagreement were as follows: Neck 24 % of all disagreements, sacrum 21 %, tail 20 %, withers 17 %, head 12 %, T13 4 % and Muzzle 2 % of all disagreements.

3.2.1. Using keypoint distance to the brush

The method to detect brushing region based on keypoint location yielded variable accuracy depending on the segment and camera placement. Macro-average accuracy reached 73 % for the side-view camera. Brushing at the tail was the most accurately detected with precision and recall of 0.99 and 0.97 respectively. Precision and recall for each body region varied with camera angle. The top-down camera had the most difficulty with the withers region (0.43 and 0.58 for precision and recall respectively) while the side view with the neck (0.6 and 0.64 respectively). Overall, accuracy was lowest from the high-angle camera, reaching only 35 %. The high angle had a good precision for rear brushing at the tail (0.94) but a high number of false negatives (recall of 0.44).

3.2.2. Using brush projection onto spline

Table 1 shows precision, recall and macro average accuracy for the spline method, which increased agreement by 15 %, compared with the keypoint distance method, resulting in an 88 % agreement. The withers location remains the one with lowest accuracy. To understand the nature of misclassifications, we drew a confusion matrix (Fig. 5).

We notice that misclassifications predominantly occurred between adjacent anatomical locations. This suggests that while the exact location was sometimes missed, the method generally identified the brush in the correct anatomical region. While there might be a failure in representing the location of the brush with sufficient accuracy, this granularity may not be required depending on the behaviour being monitored. For instance, the confusion matrix (Fig. 5) shows that misclassifications often happen in adjacent areas, particularly the head and the neck. If frontal brushing is of interest, regardless of specific segment, neck and head could be grouped for a marginal increase in agreement of 3 %, resulting in a 91 % agreement. Likewise, the tail and sacrum regions could be grouped to represent rear brushing, which would increase agreement by 1 % (92 %).

Among the groupings, combining the tail and sacrum into a single "rear" category and merging the head and neck into a "front" category yielded the most significant increases in accuracy. Grouping the withers and T13 was considered less informative due to the anatomical distance between these regions.

3.3. Descriptive statistics of brushing bouts based on human annotation and automated detection

Two hundred and twenty two (222) separate occurrences of brush-use bouts by at least one cow were recorded over a period of 25:30 h, or an average of 8,7 bouts per hour. This is considering that two separate brush interactions by the same cow separated by less than 25 s are considered the same bout. The interval of 25 s was obtained by identifying an elbow on the distribution histogram of pauses (Fig. 3 A.).

Fig. 6 shows the brushing location based on either the visual observations or the spline projection method using 2D pose. A clear

Table 1
Precision, recall and macro average accuracy for brush location detection by pose estimation. Brushing location was determined by projecting the centre of brush bounding box onto a spline fitted through the keypoints.

Camera		Top-down		Side view		High angle	
Brush location	Occurrences	precision	recall	precision	recall	precision	recall
Head	4457	0.85	0.88	0.89	0.95	0.82	0.53
Neck	2755	0.76	0.78	0.77	0.85	0.46	0.56
Withers	1637	0.76	0.75	0.80	0.80	0.40	0.21
T13	1669	0.86	0.72	0.86	0.83	0.34	0.13
Sacrum	2762	0.78	0.85	0.93	0.87	0.74	0.47
Tail	20438	0.98	0.97	0.98	0.98	0.81	0.96
Accuracy		82 %		88 %		59 %	

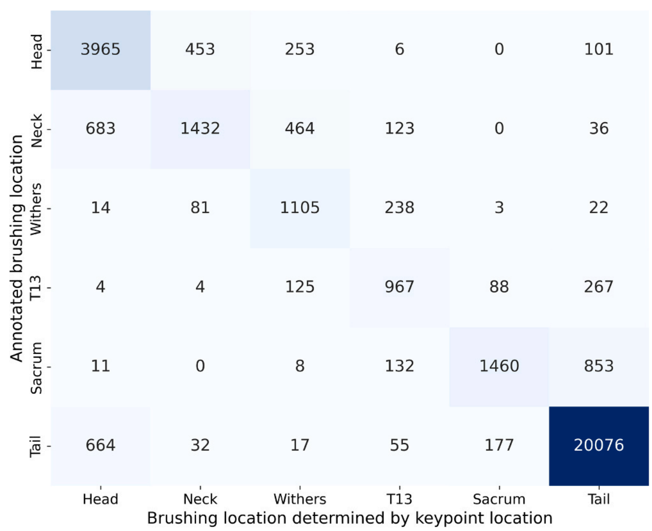

Fig. 5. Confusion matrix comparing annotated brushing locations to those determined by keypoint-to-brush proximity as projected on a spline approximating the cow's backline.

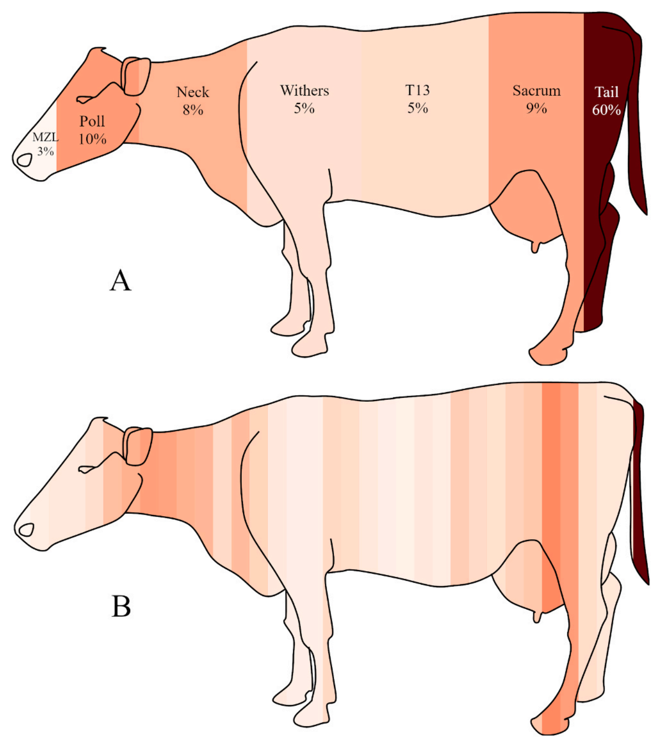

Fig. 6. Brushing location frequencies (colour scale for log-10 relative frequencies). Plot A shows brushing frequencies in each body region as annotated by a human observer. Plot B shows brush distribution across the body based on the frequency of brush positions on the body.

propensity to brush the rear, beyond the tail, and to some extent on the sacrum was observed. The neck and the back of the head seem to be another region where cows tod to brush.

4. Discussion

This study explored if an analytical approach using pose estimation can be implemented to determine the body segment being brushed by a mechanical brush in a group of loose-housed dairy cattle. Two supervised methods to predict the occurrence of brushing were also evaluated.

4.1. *Predicting brushing bouts using machine learning on pose estimation data*

Predicting brush use with machine learning yielded an accuracy of 86.3 % and F1 score of 0.81. This metric is accurate in terms of animal behaviour benchmarks; as a comparison, detection of brush–use by RFID had an F1 score of 0.78 (Sadrzadeh et al., 2024). This metric per-se is encouraging in terms of development of automated methods. However, in this two-step approach (detecting brush-use then anatomical location) any errors in the initial brushing detection would propagate to the subsequent classification step, potentially compromising overall accuracy.

The recording duration used in this study (25:30 h) may be too small to capture the variability in brush-use patterns with a neural network. This in practice means that the model was exposed to few different patterns (Zheng et al., 2014), which also likely showed temporal correlations. The LSTM attempted to classify rolling windows for brush-use status, including a temporal memory of 16 windows (00:01:20). A memory of 32 windows, as well as a bi-directional layer of 16 windows and two combinations of dense layers were also tried and yielded similar accuracies and convergence patterns. The random forest yielded a comparable accuracy to the LSTM but does not incorporate the continuous time dimension. It is likely that other features than the keypoints distance, for instance their displacement, would be more informative and produce a prediction that is sufficiently robust.

The accuracy, driven by a high number of false positives was insufficient to draw reliable conclusions on the animals and should be improved for practical applications. Since the focus of this study was on the segmentation of the brush-use location, it was preferred to select sequences based on the ground truth, to avoid transferring errors between steps of the analysis.

A much simpler approach, which was not explored, would be to log when the brush is activated. This has previously been achieved by either adapting current transducers (Falk et al., 2018) or by utilizing data by integrated light sensors used to detect tilting of the brush arm to activate rotation of the brush (Sadrzadeh et al., 2024).

4.2. Determining the body segment being brushed using pose estimation

The detection of anatomical locations on the other hand yielded encouraging accuracy of up to 87 % hinting that pose estimation is an effective tool for brush-use patterns. Considering that the observer had an agreement of 90 % when re-evaluating the video, the detection is nearly as consistent with the observer as they are with themselves. To the best of our knowledge there are no previous studies focusing on automated detection on body segment use of mechanical brushes, or other comparable resources. There is therefore a need to further explore methods in this area, to provide accurate and effective applications.

The detection of body segment being brushed is not trained on annotated data, meaning that it does not incorporate the discrepancies of the observation in its development. Rather, it leverages the high interpretability of pose estimation to recreate the annotations based on the relative location of anatomical features to the brush. This approach also means that the detection per-se is not influenced by potential observer bias and is not over-fit to the observer, thus transposable to other contexts. We consider with this performance and method that the model can be considered reliable for practical applications and that it should have high external validity.

There remains a level of disagreement, especially on the front half of the cow using the segmentation proposed here. Possible reasons for misclassification include keypoint jitter and the difficulty in representing the edges of a round object (the brush) when predicted with a square bounding box. That is, when the cow is brushing at the neck for instance (an anatomical location often misclassified), if the cow is slightly wrapping around the brush to maximise brushing area, the resulting brush keypoint at that camera angle might be closest to the withers. Moreover, the brush bounding box depends on the angle of view from the camera and its centre may not represent the actual centre of the brush in real space. A possible way of addressing this would be to use explicit body segmentation. This family of models predicts a mask of exact the pixels pertaining to each anatomical region (or object). From the predicted masks for each anatomical region, we could then compute border regions with the brush mask. Body segmentation of cows has been successfully developed by Jia et al. (2021) and could be transposed to brush-use detection.

2D pose estimation has inherent limitations when determining locations in space. 2D pose estimation operates within a two-dimensional grid, projecting the spatial configuration of a cow's body onto the image plane. This inherently introduces several issues, namely depth ambiguity, occlusions and perspective scaling (Ben Gamra and Akhloufi, 2021); firstly, in a two-dimensional image, depth information (i.e., the third dimension) is reduced. Consequently, two points equidistant in the image plane may not be equidistant in three-dimensional space. Secondly, perspective scale means that objects closer to the camera appear larger and that two equidistant points on the image are thus not equidistant in a 2D grid projected over the scene. This in turn affects the comparability of the results, between different anatomical features, and between different studies. Finally, the 2D representation is sensitive to occlusions, especially from the side view, for example when another cow is passing in front of the one using the brush. In our study, the initial results (Table 1) have shown that accuracy is sensitive to camera placement. A way to address these limitations is multi-view pose-estimation, generating a representation of the keypoints in 3D (Ben Gamra and Akhloufi, 2021).

Ultimately, the choice of method is dependent on the goals and means of a behavioural study. Visual observations are time consuming but may get closer to the true value of brush locations. Detecting brush locations using pose estimation on the other hand is more prone to errors but allows describing brush use patterns on a continuum rather than predefined segments. Fig. 6 shows that depending on the method, the way the behaviour is approached and measured changes. This has implications for the conclusions one might take. Researchers should take into account how the method informs on the behaviour when designing a study, but also how the measure of the behaviour is affected by the choice of sensor.

4.3. Defining brushing behaviour in dairy cows from an automated perspective

To develop a robust system with an accurate understanding of brush use behaviour in dairy cows, further exploration and clear definition of both body segments involved and bout duration are essential. While this study did not aim to define bout length or body segments, refining these parameters remains critical for improving classification accuracy and the overall reliability of automated monitoring systems.

Several definitions of anatomical regions for brush use have been suggested (Burton and Blackie, 2024; DeVries et al., 2007; Toaff-Rosenstein et al., 2017). In this study, the definitions of segments were based on the keypoints that existed in the pose estimator. The reasoning was to develop an ethogram suitable for both human annotation and at the same time adapt the annotations to features that are "machine learnable" (Brouwers et al., 2023). Future studies are needed to explore and define biologically important regions for the cows and take into account successive brushing locations.

Many behaviours can be described in terms of bouts, where the performance of the behaviour is bound in time. However, behaviours may be interrupted by brief pauses between successive performances of the behaviour, by pauses within a single bout, or by longer time periods between bouts (Horvath and Miller-Cushon, 2019). In this study, a brushing bout for dairy cows using a mechanical cow brush was defined as successive physical interactions with the brush separated by less than 25 s. The cut-off at 25 s was determined by analysing the distribution of interruption durations. Previous definitions of brushing and grooming bouts have either been defined without clear justification; (1) 30-second criterion for allogrooming bouts in dairy cows (Val-Laillet et al., 2009), or (2) based on complex bout criterion calculations with 125.9 s for brush use in dairy calves (Horvath and Miller-Cushon, 2019). At the same time, it has been highlighted that these pauses are influenced by interaction with other cows and environmental factors (Foris et al., 2023), as well as proximity to other resources (Mandel et al., 2013). Therefore, context-specific definitions might be important to evaluate. Defining bouts was outside the scope of this study but should be considered and evaluated when developing an automatic monitoring system.

5. Conclusions

To our knowledge, this is the first study exploring the automation of brush use patterns assessment in dairy cows through body segmentation, enabling a more detailed assessment of brush usage. The method showed sufficient agreement with the observer (even more so when grouping close body segments together) and could be applied on longer timescales to increase our knowledge on brush-use behaviour. The approach is constrained by the limitations of 2D pose estimation and proved sensitive to camera placement. Further work with sensor fusion could address these limitations. Particularly, fusing 2D detections from all cameras into a 3D pose – which is invariant to camera location (Ma et al., 2021) – could increase robustness, while RTLS could add individual recognition.

Funding

This project was funded by Formas – a Swedish Research Council for Sustainable Development, Stockholm, Sweden (ID: 2021–02254).

CRediT authorship contribution statement

Adrien Kroese: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Per Peetz Nielsen:** Writing – review & editing, Funding acquisition, Conceptualization. **Moudud Alam:** Writing – review & editing, Supervision, Methodology. **Nils Fall:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Lena-Mari Tamminen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Niclas Högberg:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **David Berthet:** Writing – review & editing, Software, Resources, Methodology, Data curation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4o in order to rephrase paragraphs and improve readability as well as to generate code for the analyses. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sony Nordic provided the technology to generate the 2D pose. Implementation, formal analysis and presentation of the results were decided by researchers at the Swedish University of Agricultural Sciences.

Acknowledgements

The authors would like to thank the staff of the Swedish Livestock Research Centre for their help and Sony for the extensive collaboration.

References

Ben Gamra, M., Akhloufi, M.A., 2021. A review of deep learning techniques for 2D and 3D human pose estimation. Image Vis. Comput. 114, 104282. https://doi.org/10.1016/j.imavis.2021.104282.

Brouwers, S.P., Simmler, M., Savary, P., Scriba, M.F., 2023. Towards a novel method for detecting atypical lying down and standing up behaviors in dairy cows using accelerometers and machine learning. Smart Agric. Technol. 4, 100199. https://doi.org/10.1016/j.atech.2023.100199.

Burton, Y.I., Blackie, N., 2024. Impact of lameness on brush use in a loose-housed dairy system. Ruminants 4, 375–386. https://doi.org/10.3390/ruminants4030027.

Chen, C., Zhu, W., Norton, T., 2021. Behaviour recognition of pigs and cattle: journey from computer vision to deep learning. Comput. Electron. Agric. 187, 106255. https://doi.org/10.1016/j.compag.2021.106255.

Chollet, F., 2015. Others. In: Keras [Internet], 2015. GitHub. Available from: ⟨https://github.com/fchollet/keras⟩.

De Oliveira, D., Keeling, L.J., 2018. Routine activities and emotion in the life of dairy cows: integrating body language into an affective state framework. PLOS ONE 13, e0195674. https://doi.org/10.1371/journal.pone.0195674.

DeVries, T.J., Vankova, M., Veira, D.M., von Keyserlingk, M. a G., 2007. Short communication: usage of mechanical brushes by lactating dairy cows. J. Dairy Sci. 90, 2241–2245. https://doi.org/10.3168/jds.2006-648.

EFSA Panel on Animal Health and Animal Welfare, Nielsen, S.S., Alvarez, J., Bicout, D.J., Calistri, P., Canali, E., Drewe, J.A., Garin-Bastuji, B., Gonzales Rojas, J.L., Gortázar Schmidt, C., Herskin, M., Michel, V., Miranda Chueca, M.Á., Padalino, B., Roberts, H. C., Spoolder, H., Stahl, K., Velarde, A., Viltrop, A., De Boyer Des Roches, A., Jensen, M.B., Mee, J., Green, M., Thulke, H.-H., Bailly-Caumette, E., Candiani, D., Lima, E., Van der Stede, Y., Winckler, C., 2023. Welfare of dairy cows. EFSA Journal. 21, e07993. doi:10.2903/j.efsa.2023.7993.

Falk, M.L., Cantor, M.C., Costa, J., Hayes, M., Jackson, J., 2018. Validation of radio frequency identification with a current transducer to quantify the use of an automatic grooming brush in pre-weaned dairy calves. Presented at the 10th International Livestock Environment Symposium (ILES X). ASABE, St. Joseph, MI, p. 1. https://doi.org/10.13031/iles.18-107.

Fernandes, A.F.A., Dórea, J.R.R., Rosa, G.J., de, M., 2020. Image analysis and computer vision applications in animal sciences: an overview. Front. Vet. Sci. 7. https://doi.org/10.3389/fvets.2020.551269.

Foris, B., Sadrzadeh, N., Krahn, J., Weary, D.M., von Keyserlingk, M.A.G., 2023. The effect of placement and group size on the use of an automated brush by groups of lactating dairy cattle. Animals 13, 760. https://doi.org/10.3390/ani13040760.

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. YOLOX: exceeding YOLO series in 2021 (version 2). arXiv. https://doi.org/10.48550/ARXIV.2107.08430.

Gong, C., Zhang, Y., Wei, Y., Du, X., Su, L., Weng, Z., 2022. Multicow pose estimation based on keypoint extraction. PLOS ONE 17 (6), e0269259. https://doi.org/10.1371/journal.pone.0269259.

Horvath, K.C., Miller-Cushon, E.K., 2019. Characterizing grooming behavior patterns and the influence of brush access on the behavior of group-housed dairy calves. J. Dairy Sci. 102, 3421–3430.

Jia, N., Kootstra, G., Koerkamp, P.G., Shi, Z., Du, S., 2021. Segmentation of body parts of cows in RGB-depth images based on template matching. Comput. Electron. Agric. 180, 105897. https://doi.org/10.1016/j.compag.2020.105897.

Keeling, L.J., Winckler, C., Hintze, S., Forkman, B., 2021. Towards a positive welfare protocol for cattle: a critical review of indicators and suggestion of how we might proceed. Front. Anim. Sci. 2. https://doi.org/10.3389/fanim.2021.753080.

Kroese, A., Alam, M., Hernlund, E., Berthet, D., Tamminen, L.-M., Fall, N., Högberg, N., 2024. 3D pose estimation to detect posture transition in free-stall housed dairy cows. J. Dairy Sci. https://doi.org/10.3168/jds.2023-24427.

Lecorps, B., Féron, C., 2015. Correlates between ear postures and emotional reactivity in a wild type mouse species. Behav. Process. 120, 25–29. https://doi.org/10.1016/j.beproc.2015.08.002.

Li, X., Cai, C., Zhang, R., Ju, L., He, J., 2019. Deep cascaded convolutional models for cattle pose estimation. Comput. Electron. Agric. 164, 104885. https://doi.org/10.1016/j.compag.2019.104885.

Littin, K., Acevedo, A., Browne, W., Edgar, J., Mendl, M., Owen, D., Sherwin, C., Würbel, H., Nicol, C., 2008. Towards humane end points: behavioural changes precede clinical signs of disease in a Huntington's disease model. Proc. R. Soc. B Biol. Sci. 275, 1865–1874. https://doi.org/10.1098/rspb.2008.0388.

Ma, H., Chen, L., Kong, D., Wang, Z., Liu, X., Tang, H., Yan, X., Xie, Y., Lin, S.-Y., Xie, X., 2021. TransFusion: Cross-view. Fusion Transform. 3D Hum. Pose Estim.. https://doi.org/10.48550/arXiv.2110.09554.

Mandel, R., Whay, H.R., Nicol, C.J., Klement, E., 2013. The effect of food location, heat load, and intrusive medical procedures on brushing activity in dairy cows. J. Dairy Sci. 96, 6506–6513. https://doi.org/10.3168/jds.2013-6941.

Mandel, R., Nicol, C.J., Whay, H.R., Klement, E., 2017. *Short communication:* detection and monitoring of metritis in dairy cows using an automated grooming device. J. Dairy Sci. 100, 5724–5728. https://doi.org/10.3168/jds.2016-12201.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., Bethge, M., 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. 21, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.

McConnachie, E., Smid, A.M.C., Thompson, A.J., Weary, D.M., Gaworski, M.A., von Keyserlingk, M.A.G., 2018. Cows are highly motivated to access a grooming substrate. Biol. Lett. 14, 20180303. https://doi.org/10.1098/rsbl.2018.0303.

McFarland, D., 1999. Animal behaviour: psychobiology, ethology, and evolution. 3rd ed. ed. Longman, Harlow, England.

Meunier, B., Pradel, P., Sloth, K.H., Cirié, C., Delval, E., Mialon, M.M., Veissier, I., 2018. Image analysis to refine measurements of dairy cow behaviour from a real-time location system. Biosyst. Eng. Adv. Eng. Sens. Based Monit. Manag. Syst. Precis. Livest. Farming 173, 32–44. https://doi.org/10.1016/j.biosystemseng.2017.08.019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.-H., Murthy, M., Shaevitz, J.W., 2019. Fast animal pose estimation using deep neural networks. Nat. Methods 16, 117–125. https://doi.org/10.1038/s41592-018-0234-5.

Russello, H., van der Tol, R., Kootstra, G., 2022. T-LEAP: occlusion-robust pose estimation of walking cows using temporal information. Comput. Electron. Agric. 192, 106559. https://doi.org/10.1016/j.compag.2021.106559.

Sadrzadeh, N., Foris, B., Krahn, J., Keyserlingk, M.A.G. von, Weary, D.M., 2024. Automated monitoring of brush use in dairy cattle. PLOS ONE 19, e0305671. https://doi.org/10.1371/journal.pone.0305671.

Schmied, C., Waiblinger, S., Scharl, T., Leisch, F., Boivin, X., 2008. Stroking of different body regions by a human: effects on behaviour and heart rate of dairy cows. Appl. Anim. Behav. Sci. 109, 25–38. https://doi.org/10.1016/j.applanim.2007.01.013.

Skånberg, L., Gauffin, O., De Oliveira, D., Norling, Y., Keeling, L., 2017. The positive type of brushing. Presente 2nd Anim. Welf. Sci. Symp 2122 August 2017 Swed. Univ. Agric. Sci. Ultuna.

Szeliski, R., 2011. Computer vision: algorithms and applications. Texts in Computer Science. Springer, London. https://doi.org/10.1007/978-1-84882-935-0.

Toaff-Rosenstein, R.L., Velez, M., Tucker, C.B., 2017. *Technical note:* use of an automated grooming brush by heifers and potential for radiofrequency identification-based measurements of this behavior. J. Dairy Sci. 100, 8430–8437. https://doi.org/10.3168/jds.2017-12984.

Val-Laillet, D., Guesdon, V., von Keyserlingk, M.A.G., de Passillé, A.M., Rushen, J., 2009. Allogrooming in cattle: relationships between social preferences, feeding displacements and social dominance. Appl. Anim. Behav. Sci. 116, 141–149. https://doi.org/10.1016/j.applanim.2008.08.005.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B., 2019. Deep high-resolution representation learning for visual recognition (version 2). arXiv. https://doi.org/10.48550/ARXIV.1908.07919.

Weary, D.M., Huzzey, J.M., von Keyserlingk, M.A.G., 2009. Board-invited review: using behavior to predict and identify ill health in animals1. J. Anim. Sci. 87, 770–777. https://doi.org/10.2527/jas.2008-1297.

Weigele, H.C., Gygax, L., Steiner, A., Wechsler, B., Burla, J.-B., 2018. Moderate lameness leads to marked behavioral changes in dairy cows. J. Dairy Sci. 101, 2370–2382. https://doi.org/10.3168/jds.2017-13120.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., Zhao, J.L., 2014. Time series classification using multi-channels deep convolutional neural networks. In: Li, F., Li, G., Hwang, S., Yao, B., Zhang, Z. (Eds.), Web-Age Information Management. Springer International Publishing, Cham, pp. 298–310. https://doi.org/10.1007/978-3-319-08010-9_33.