JOURNAL OF
EMERGING INVESTIGATORS

Article

Recognition of animal body parts via supervised learning

Betina Kreiman[1], Jaeson Jang[2,3], Lakshmi Narasimhan Govindarajan[2,3], Thomas Serre[2,3]

[1] Newton North High School, Newton, Massachusetts

[2] Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI

[3] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI

SUMMARY

The application of machine learning techniques has facilitated the automatic annotation of behavior in video sequences, offering a promising approach for ethological studies by reducing the manual effort required for annotating each video frame. Nevertheless, before solely relying on machine-generated annotations, it is essential to evaluate the accuracy of these annotations to ensure their reliability and applicability. While it is conventionally accepted that there cannot be a perfect annotation, the degree of error associated with machine-generated annotations should be commensurate with the error between different human annotators. We hypothesized that machine learning supervised with adequate human annotations would be able to accurately predict body parts from video sequences. Here, we conducted a comparative analysis of the quality of annotations generated by humans and machines for the body parts of sheep during treadmill walking. For human annotation, two annotators manually labeled six body parts of sheep in 300 frames. To generate machine annotations, we employed the state-of-the-art pose-estimating library, DeepLabCut, which was trained using the frames annotated by human annotators. As expected, the human annotations demonstrated high consistency between annotators. Notably, the machine learning algorithm also generated accurate predictions, with errors comparable to those between humans. We also observed that abnormal annotations with a high error could be revised by introducing Kalman Filtering, which interpolates the trajectory of body parts over the time series, enhancing robustness. Our results suggest that conventional transfer learning methods can generate behavior annotations as accurate as those made by humans, presenting great potential for further research.

INTRODUCTION

Traditionally, the assessment of animal behavior has relied on vast amounts of human labor by manual scrutiny of animals in natural habitats or video sequences. For example, in a memory study conducted in 1982, human annotators manually recorded the time rats took to find a fixed location (1). However, such manual annotations are error-prone, difficult to reproduce, and remain a primary bottleneck in the systematic analysis of animal behavioral data (2). Recently, a combination of novel machine learning algorithms, increased computational power, and inexpensive data storage have promised to revolutionize ethological studies by training computers to automatically evaluate animal behavior (2-4).

A prominent example is DeepLabCut, a software developed for pose estimation based on transfer learning with deep neural networks (2, 3). This is a markerless approach because it does not require introducing any apparatus or physical labels on the animals themselves, which may interfere with their natural behavior. This program was developed as a substitute for labor-intensive human annotation of animal body parts in video frames, so manual labor is restricted to annotating minimal training data, typically 50-200 video frames. The DeepLabCut model is then trained through supervised learning with these training data, enabling it to predict the location of body parts in all other frames automatically. Supervised learning refers to a machine learning procedure where neural network weights are adjusted according to labels in the training data. As a result, DeepLabCut could provide a robust and accurate annotation of body parts as human labeling for mice, monkeys, and macaques (2, 3, 5, 6). Additionally, only a small number of parameters are needed to obtain accurate results. These results suggest the promising potential of using deep learning algorithms for animal behavioral analysis, but it is essential to evaluate the accuracy of such machine-generated annotations when the algorithm is exposed to new data.

Sheep are a commonly studied species in spinal cord injury research due to their large spinal cord size, which serves as a useful anatomical surrogate for humans (7). In particular, investigating the gait cycle of sheep walking on a treadmill is a widespread method that can be used to examine the impact of electrical stimulation on kinematics. Electrical stimulation can also be used as a therapy to improve locomotor function, as in the case of Parkinson's patients. Automatic evaluation of animal behavior can also help in the case of animals that undergo surgery for diagnoses, potential treatments, and recovery assessment.

The traditional and manual experimental approach requires an expensive motion-capturing system to automatically obtain the coordinates of each body part from video frames for reconstructing the gait cycle. While it is technically possible to annotate body parts manually for each frame without such a motion-capturing system, this would require a significant amount of human labor (2, 4). For example, continuous manual annotation of 1 minute of the gait cycle with 6 body parts would require working on 1,200 frames at a recording frequency of 20Hz. Such annotations require about 5 hours of human work. In contrast, after training, machine learning algorithms would take a fraction of a second for the same number of labels.

DOI: https://doi.org/10.59720/23-109

In this study, we evaluated the effectiveness of using DeepLabCut for automated annotation of the body parts of sheep. We tested our hypothesis that supervised learning would suffice to extract body parts in moving animals from video data. The first step was to obtain human annotations for the DeepLabCut models to be trained with (2, 3). Annotating individual video frames takes humans approximately 5 hours per minute of video. Given the large cost of manual annotation, we aimed to assess the minimum amount of annotations that would suffice to train robust machine learning algorithms. Additionally, we wanted to compare differences between humans and also differences between humans and machines. Thus, we decided to consider two independent human annotators. We investigated whether the machine-generated annotations of each body part could match the range of locations annotated by multiple human annotators, and thus provide a reliable annotation for research purposes. Our supervised machine learning algorithm generated accurate predictions, with errors comparable to those observed between human annotators. We also implemented a method to overcome errors due to occlusion. Our results suggest that conventional transfer learning methods can be effective in generating reliable annotations for sheep walking on a treadmill, and thus can be applied in spinal cord injury research with great potential. In the future, similar approaches might be applied to human behavior and motor injury recovery.

RESULTS

Annotations by different humans were consistent with each other

In this study, we aimed to compare the consistency between annotations generated by human annotators and the DeepLabCut toolbox in the analysis of sheep treadmill videos from three different angles (**Figure 1a**). The human annotations were obtained by asking two independent human annotators to label the locations of six body parts (foot, knee, and thigh of left and right legs) in 150 video frames. These frames were split into training (80%) and testing (20%) datasets, with the DeepLabCut toolbox being trained on the former and used to predict body part locations in the latter (**Figure 1b**).

To assess the consistency between human annotations, we calculated the Euclidean distance between labels for the same body part in each frame (**Figure 2**). In some frames, the annotations made by both human annotators almost completely overlapped within a few pixels (**Figure 2a**, top). In most frames, there was a noticeable discrepancy between the two annotations (**Figure 2a**, bottom). Across all body parts and recording angles, the average distance between the two human annotations was 21.3 ± 18.6 pixels (**Figure 2b**; frame size = 880 x 620 pixels). This difference is well below what one would expect from random annotations, which would lead to an average distance of 390.3 ± 194.3 pixels ($p < 10^{-20}$). The average distance under the null hypothesis of random annotations was obtained by simulating random predictions the same number of times as human annotations (150 frames x 6 body parts x 2 annotators = 1800 times). In addition, the average error within each condition was smaller than the average distance between body parts (knee-foot: 149.2 pixels, thigh-knee: 184.2 pixels, thigh-foot: 333.4 pixels), implying a consistent level of human annotation (**Figure 2c**).

Computer predictions matched human annotations in test frames

To validate the consistency of our model annotations, an additional set of 20 frames for each camera angle was selected at random from video recordings and was annotated by both human annotators and the trained models (**Figure 3a**). It is noteworthy that these frames were not utilized during the training phase of the DeepLabCut models, providing an opportunity for cross-validation. The Euclidean distance between computer-generated annotations and human annotations (15.3 ± 25.0 pixels, average of the distribution) was

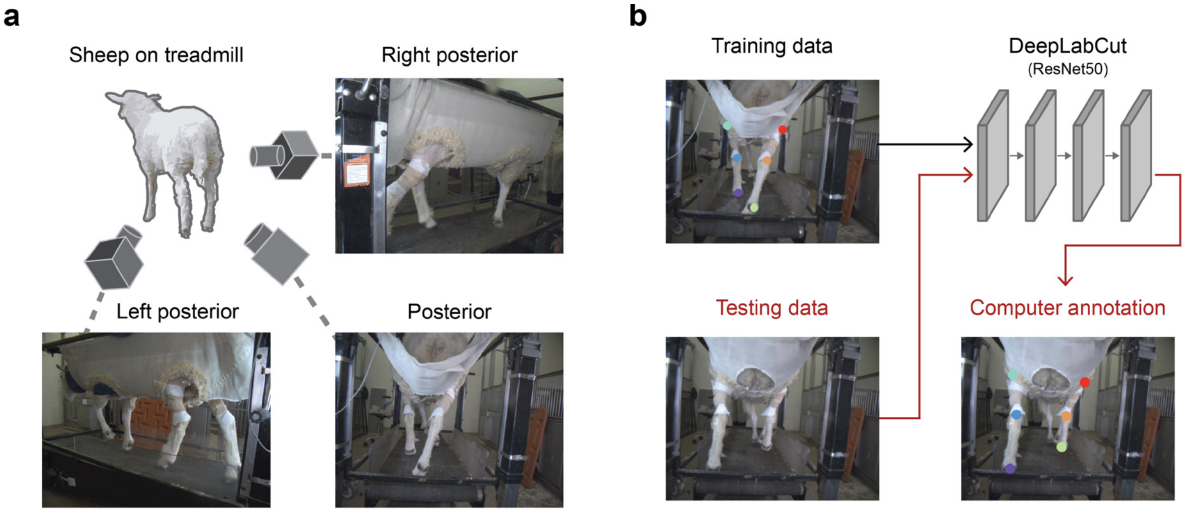

Figure 1: Schematic of the overall research approach. a) Example frames showing the three different camera angles to visualize the right posterior, left posterior, and posterior of the animal. These frames were used to train DeepLabCut. **b)** Top image is an example frame annotated by a researcher. The annotator labeled 6 locations: left foot, left knee, left thigh, right foot, right knee, and right thigh. These labeled locations were fed to the DeepLabCut neural network (ResNet) for supervised training. The ResNet architecture is schematically illustrated here by a series of rectangular shapes that refer to each layer of the neural network. Only four of the 50 layers are shown here for simplicity. Bottom images show the predictions made by DeepLabCut for each body part on test frames that were not used for training.

DOI: https://doi.org/10.59720/23-109

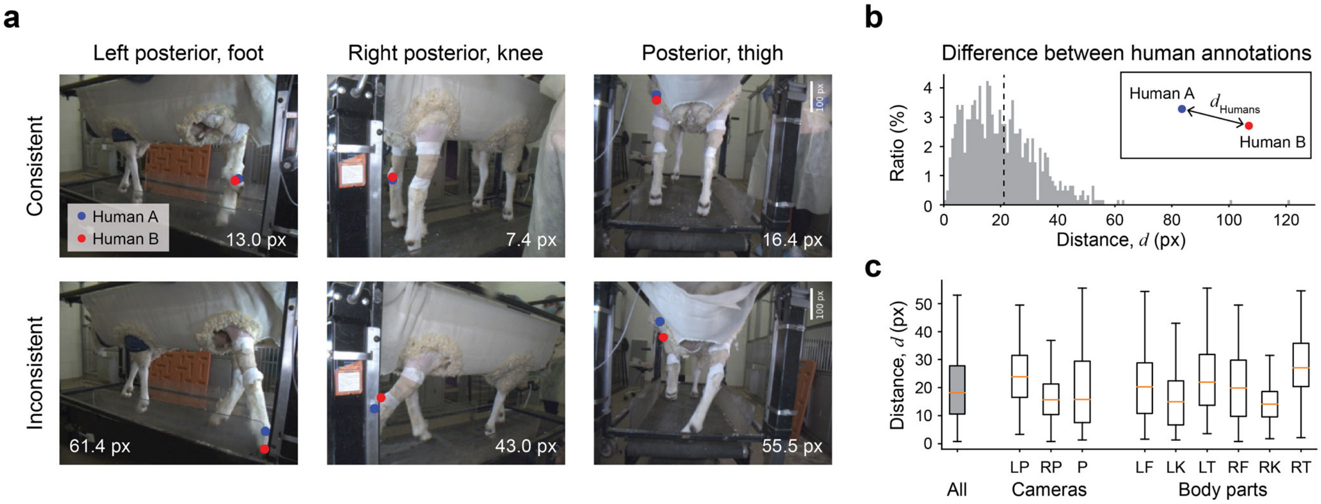

Figure 2: Consistency of human annotations. a) Top images are representative example frames showing similar annotations between both human annotators. Bottom images are example frames showing more inconsistent annotations between the human annotators. The average Euclidean distance (px) between annotations is written in each frame. Similar results were obtained for body parts not shown. In the case of consistent annotations (top), the two location labels tend to overlap and may be hard to distinguish. **b)** Distribution of distances between human annotations combined over all body parts and camera angles. The average distance between human annotations is 21.3 ± 18.6 pixels (denoted by the dashed line). Total number of annotations is 150 x 6 = 900. **c)** Distribution of Euclidean distances when separated into camera angles and into each body part. Here and subsequently: LP = left posterior angle, RP = right posterior angle, P = posterior angle, LF = left foot, LK = left knee, LT = left thigh, RF = right foot, RK = right knee, RT = right thigh. There was a small amount of variation in the degree of consistency between annotators across different camera angles but this did not reach statistical significance ($p > 0.01$, One-Way Anova test). There was a statistically significant difference in the degree of consistency across different body parts ($p < 10^{-5}$, One-Way Anova test).

similar to the distance between the two human annotations (12.7 ± 9.6 pixels) for the validation set (**Figure 3b**). There was no statistically significant difference between these two distributions across all 6 body parts ($p = 0.84$). These findings suggest that the models were accurate in their predictions.

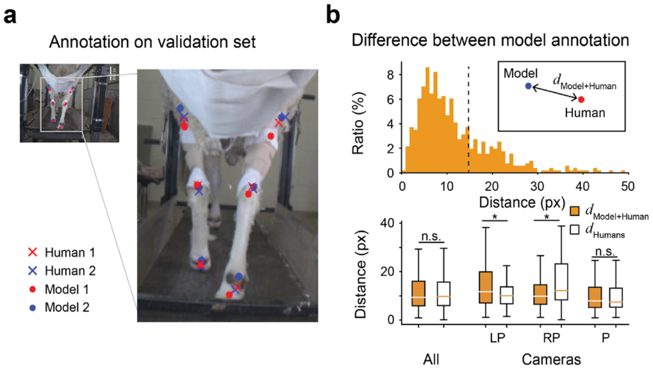

Figure 3: The neural network accurately predicted body parts for cross-validated test frames. a) An example frame in the cross-validation set with both human annotations and both model annotations. The predictions are similar. **b)** Distribution of distances between the DeepLabCut model predictions based on each of the two annotators (n = 120 frames x 6 body parts x 2 annotators). This comparison is based on test frames not used for training and validation in the prior sections. The dashed line represents the average Euclidean distance of 15.3 ± 25 pixels. The posterior view average distances are very similar ($p = 0.51$). The left and right posterior views were not as similar ($p < 10^{-16}$ and $p < 10^{-6}$, respectively), but this was most likely due to occlusion (see section, **Difficulty in predicting the location of occluded body parts**).

Computer predictions obtained from different annotators were consistent

Next, we sought to evaluate the consistency between computer annotations generated by DeepLabCut and human annotations. Specifically, we trained two separate DeepLabCut models using the annotations generated by each human annotator (**Figure 4a**). Given the different training data, it is conceivable that the two models could lead to different results. We compared the predicted locations of body parts in the testing dataset by measuring the Euclidean distance between two points. In some frames, the annotations made by both DeepLabCut models were almost completely overlapped within a few pixels while there was a difference between the two annotations in most frames (**Figure 4b**).

Across all camera angles and body parts, we found that the average distance error between the two computer models was significantly larger than the inter-human error (between models = 45.3 ± 67.6 pixels; between humans = 21.3 ± 18.6 pixels; $p < 10^{-12}$) (**Figure 4c**). Notably, the computer models showed varying levels of error across camera angles, with the lowest error observed for the posterior angle and the highest for the right posterior angle (posterior = 23.6 ± 31.4 pixels, left posterior = 47.3 ± 58.0 pixels, right posterior = 65.0 ± 92.1 pixels) (**Figure 4c**). While the difference in error between computer models for the posterior angle was not significantly different from that of humans ($p = 0.078$), we observed a larger difference in error for the side-view recordings. This was primarily due to the occlusion of body parts by the rack of the treadmill, which made it difficult for the computer models to accurately annotate these parts.

In summary, we found that DeepLabCut generated annotations that were comparably consistent with human annotations for the posterior view, where all body parts were observable in most frames. However, we observed a

DOI: https://doi.org/10.59720/23-109

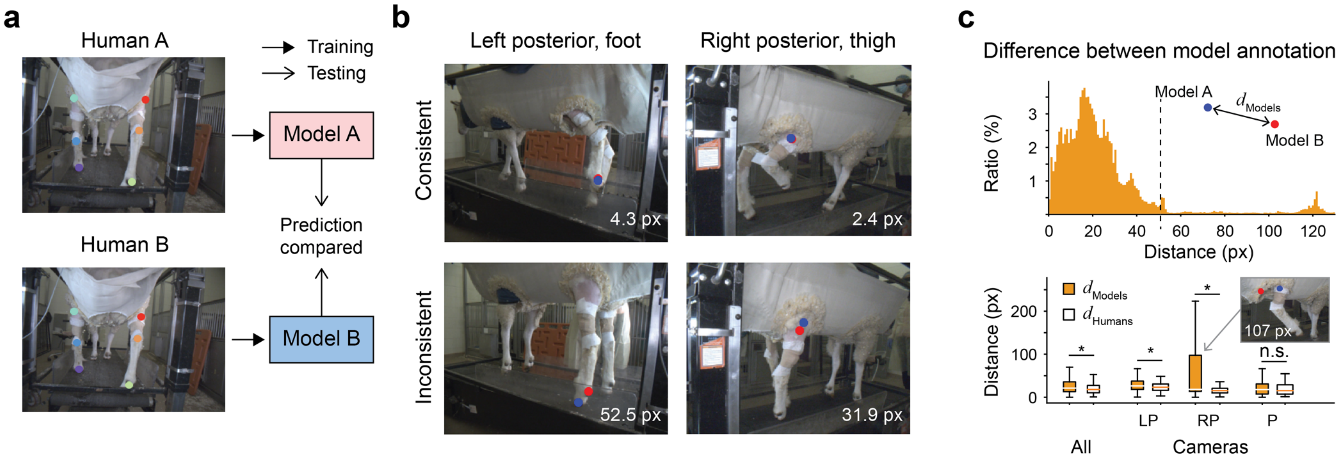

Figure 4: Consistency of model annotations. a) Two sets of human annotations were used to train two separate DeepLabCut models and their predictions were compared. **b)** Top images are representative example frames showing similar annotations between both model predictions. Bottom images are example frames showing more inconsistent annotations between the model predictions. The average Euclidean distance (px) between annotations is written in each frame. Similar results were obtained for body parts not shown. In the case of consistent annotations (top), the two location labels tend to overlap and may be hard to distinguish. **c)** Top graph shows the distribution of distances between model annotations. The x-axis denotes the Euclidean distance between the two model annotations combined over all body parts and camera angles. The average distance between model predictions is 45.3 ± 67.6 pixels (denoted by the dashed line). Total number of annotations was 5,710 frames x 6 body parts x 3 camera angles = 102,780. Bottom graph shows the distribution of Euclidean distances when separated into camera angles (left posterior, right posterior, and posterior views) compared to human annotators. Inset indicates an example frame where two models predicted relatively distant locations for the same body part.

significant difference in error between computer models and human annotators for other camera angles, suggesting the influence of occlusion on the accuracy of computer annotations.

Difficulty in predicting the location of occluded body parts

A larger error between computer models compared to that between human annotators was primarily observed from side view recordings (left and right posterior views), which frequently involve the occlusion of body parts across multiple frames (**Figure 4**). In frames where a body part was not visible, the human annotators did not label it, and it was consequently excluded from the calculation of the average Euclidean distance between human annotations. However, for such frames, the computer model predicted the location of the missing body part with a low confidence level, and these predictions were still considered in calculating the distances between computer models in the previous analysis (**Figure 5a**, yellow dots predicting the location of the right thigh). It is noteworthy that, when the error distance between model predictions exceeded 200 pixels, 77.0% and 81.1% of model predictions for the left and right posterior views exhibited a confidence level of less than 0.1 on a scale of 0 to 1. When the confidence level was less than 0.1, the annotation was not visible in the video created by DeepLabCut.

In order to address the issue of erroneous predictions for frames where specific body parts are occluded, we implemented a solution involving Kalman filtering. This technique involves smoothing the trajectory of a given body part across neighboring frames over time, thereby minimizing the impact of sporadic or anomalous predictions (10). After applying the Kalman filter, the average Euclidean distance between an annotator and the model decreased for body

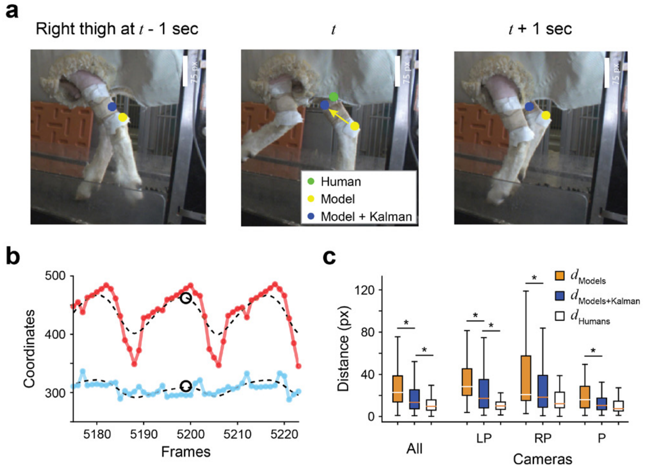

Figure 5: Kalman filtering enhances the annotation consistency. a) Frames at t - 1 second, t, and t + 1 second with model predictions of the right thigh (yellow dots) alongside revised predictions obtained through Kalman filtering of annotated locations over time (blue dots). The second image shows an example that the Kalman filtering improves the model annotation for the body part which is occluded in the frame. A green dot represents the human annotation as the ground truth. **b)** The temporal trajectory of model predictions in x- and y-coordinates over time for 2 seconds (blue and red) and the smoothed trajectory through Kalman filtering (black dashed lines). The observation covariance as a smoothing parameter was set to 10, where an observation covariance of 0 indicates no smoothing. The black circles represent the human label for one annotator at that particular frame. **c)** Distance errors between the model predictions, the smoothed model predictions through Kalman filtering, and the human annotations. Original model predictions were not very similar to smoothed predictions after Kalman filtering (all: $p < 10^{-14}$, LP: $p < 10^{-7}$, RP: $p < 10^{-2}$, and P: $p < 10^{-5}$). Human annotations were more similar to smoothed model predictions after Kalman filtering (all: $p < 10^{-8}$, LP: $p < 10^{-7}$, RP: $p < 10^{-1}$, and P: $p < 10^{-1}$).

DOI: https://doi.org/10.59720/23-109

parts that were occluded. For example, for the right posterior camera angle, we applied the Kalman filter to smooth over the annotations labeling the right thigh (**Figure 5a**). An example of the improvement from Kalman Filtering can be seen in **Figure 5b**. Following the application of Kalman filtering, the average Euclidean distance between an annotator and the model decreased by 12.9 pixels (before Kalman Filtering 127.5 ± 13.8 pixels, after Kalman Filtering 114.6 ± 25.2 pixels, 11.2% change). It is noteworthy that the distance between DeepLabCut model predictions significantly decreased across all the cameras, indicating the improvement in the consistency of model annotations (**Figure 5c**; p = $2.5\text{x}10^{-14}$, $1.4\text{x}10^{-7}$, 0.005, $8.1\text{x}10^{-5}$ for all cameras, left posterior, right posterior, posterior cameras, respectively).

DISCUSSION

Understanding continuous human and animal behavior is a gargantuan task that used to require tedious, error-prone, and laborious manual annotations from dedicated research assistants. The advent of Artificial Intelligence has provided modern tools to revolutionize neuroethology (4). Here we show that minimal human annotations can be used with DeepLabCut to automatically and precisely annotate body parts in footage of moving sheep. Given the successes of supervised machine learning in other domains ranging from clinical diagnoses to face recognition to discovering new galaxies, we hypothesized that DeepLabCut would create robust annotations, which could be more consistent, more precise, and faster than human annotations. More training data always helps in supervised learning approaches. Remarkably, even though we only annotated a total of 150 frames per person, we observed that this amount of training data was sufficient for supervised training of DeepLabCut. This amount of data is consistent with other efforts using DeepLabCut (2, 3). Across 5,710 frames, two independent DeepLabCut models yielded consistent annotations and accuracy was verified in cross-validated test data. Of note, the evaluations were made on test data that were not used during the training process.

One of the biggest challenges for DeepLabCut was occlusion. For example, the right leg could occlude the left leg in such a way that the left knee was not visible in the frame. Such occlusions led to the largest errors in localizing body parts and were typically associated with low confidence in the model's prediction. This necessitates an additional iteration of post-processing conducted by human annotators and consequently leads to a notable reduction in the overall efficiency of automatic annotation algorithms. When a body part became occluded due to movement, we found that it often resulted in a big change in the computer prediction (often an erroneous annotation of a different body part or a different leg). This suggests that in cases where a specific body part is occluded due to movement in a frame, the model generates a prediction with a low confidence level and occasionally mislabels the location, leading to a larger error between the predictions of the two models. Indeed the videos generated by DeepLabCut omit predictions with confidence levels below 0.1. As a potential solution, we implemented a Kalman filter that capitalizes on the temporal sequence to smooth the data and interpolate between frames. This procedure improved the predictions. The quality of model annotations could be improved by a straightforward temporal trajectory-based post-processing, which could be explored in further research on automatic annotation systems. Although we investigated the impact of filtering on calibrating occlusion within the 2D trajectory for each camera angle, it is noteworthy that a similar filtering process could also enhance the precision of 3D pose estimation, as it is used for converting multiple 2D trajectories into a 3D trajectory in another open-source toolkit, Anipose (8).

Digital storage space has become relatively inexpensive, and there are excellent digital cameras available. Therefore, it is relatively straightforward to collect vast amounts of video sequences documenting animal behavior. In the past, manual annotation of such sequences was a prohibitive bottleneck (2-4). The astounding success of DeepLabCut with small amounts of training data, as demonstrated here, opens the doors to new possibilities in computational ethology. In addition, human annotations require a lot of time (approximately 5 hours per minute of video), while training DeepLabCut for the entire project took about 10 hours. After training, automatic annotations of entire videos took seconds.

In the future, it would be interesting to use 3D DeepLabCut to create predictions of movement in 3D. Given the fact that we already used 3 different camera angles simultaneously recording the movements, it would be useful to implement 3D analysis to obtain better predictions of sheep movement. Another possible improvement could be to average two or more DeepLabCut models because this could lead to more accurate predictions. Ensembling models is a common approach in machine learning where one model can compensate for the failures of another model (5). It would also be interesting to assess whether the algorithm can extrapolate to different conditions such as changes in size, age, color, illumination, or walking speed. It is important to note that the results presented here are based on one sheep. It will be necessary to assess whether annotations in one animal can extrapolate to other animals in the future. There are also similar algorithms such as DeepPoseKit, and it would be interesting to compare the results across algorithms to assess which one provides more robust, accurate, and efficient predictions (9).

We found DeepLabCut software to be accurate and robust in its predictions. DeepLabCut created annotations similar to those that a human would create. The approach could therefore be applied to predict the trajectory of human limbs and for the creation of prosthetic limbs. In particular, in the context of spinal cord injuries, there is strong interest in automatic approaches to rigorously quantify behavior for the purposes of diagnoses, treatment, and recovery assessment.

MATERIALS AND METHODS

Study subjects

All study procedures were conducted with the approval of the Brown University Institutional Animal Care and Use Committee and in accordance with the National Institutes of Health Guidelines for Animal Research. One female sheep (*Ovis aries*) was used for this study. The animal was kept in a separate cage in a controlled environment on a 12 h light/dark cycle with *ad libitum* access to water and was fed twice daily.

Human annotations

The first step was to obtain human annotations for the DeepLabCut models to be trained with (2, 3). Two human

DOI: https://doi.org/10.59720/23-109

annotators separately labeled 6 body parts: left foot, left knee, left thigh, right foot, right knee, and right thigh (**Figure 1**). These annotations encompassed 50 frames for each one of 3 camera angles (**Figure 1a**). Thus, there were 150 frames to annotate with 6 body parts each. If a body part was occluded, the human annotator did not annotate it, so the final numbers were 279, 224, and 199 total annotations out of 300 per camera. The annotations were performed using DeepLabCut's labeling feature. The frames were selected at random by DeepLabCut and were sampled uniformly in order to obtain times when the sheep were moving and when they were still. Frames were extracted from videos of sheep walking on a treadmill.

The videos were then analyzed and labeled by the DeepLabCut (v.2.2.2) model (see **Evaluating the neural network** for details). Each video (3 in total) contained 5,710 frames. Each frame was 880 x 620 pixels with RGB colors. To annotate the body parts, there was minimal discussion between the two human annotators regarding part definitions. If a human annotator deemed a body part to be occluded, they did not label it, and it was not taken into account when calculating the average Euclidean distance between human annotations. To cross-validate and test the neural networks, each human annotator labeled 20 more frames per camera angle (60 in total), which were again randomly selected by DeepLabCut.

Training the neural network

Each of the two sets of human annotations was used to train a separate DeepLabCut model (2, 3) (**Figure 4a**). Each model was trained for 1 million epochs. Separate annotations were used for training and testing the DeepLabCut models. We used default DeepLabCut parameters for training: batch size = 8, training fraction = 0.95, network type = resnet50, augment = none, autotune = false, keepdeconvweights = true. An NVIDIA TITAN X GPU was used to train the DeepLabCut algorithm.

Evaluating the neural network

The trained DeepLabCut model generated predictions and likelihoods for each body part in each of the 5,710 frames per camera angle (**Figure 1b**). We evaluated the neural network by comparing the DeepLabCut model predictions against the human annotations on independent test data for a subset of 60 frames (20 per camera angle). In the labeled videos that DeepLabCut created, if the model's likelihood for a certain point was below 0.1, the annotation was not shown in the video. However, points with low likelihoods were still taken into account in computing the distances between network-created annotations.

Kalman filtering

In order to fix potential errors due to occlusion, we smoothed the data using Kalman Filtering (10). The goal of the filtering was to remove network predicted outliers. The smoothing parameter was the observation covariance and set to 10, determined by a manual inspection. A higher observation covariance signifies that we had less confidence in the model's predictions, creating a smoother line between points as shown in **Figure 5b**.

The Kalman Filter smooths data according to the previous data point and current predictions from the model. The filter first uses the model's prediction, updates it using the Kalman Gain and the State Updates Equation, and then predicts the new output (10). This was done for each computer-created prediction (6 body parts x 5,710 frames per camera angle).

General statistics and analyses methods

To compare different annotators, we computed the Euclidean distance between their labels for the same body parts for model predictions and for human annotations. Through the paper we used the Wilcoxon rank-sum test to make comparisons between distributions. A difference between two distributions was considered to be statistically significant if the Wilcoxon rank-sum test yielded a p value less than 0.01. These comparisons were made with human-annotated images that had not been used to train the DeepLabCut models. We compared the predictions made by models based on different annotations. These comparisons consisted of all 5,710 frames from each of the 3 videos used. All of the source code is publicly available through the following link: https://github.com/betinatkreiman/DeepLabCut_Manuscript.

ACKNOWLEDGMENTS

We would like to thank David Borton, Jonathan Calvert, and Samuel Parker from the Borton Lab for their contributions in data collection and facilitating access to the datasets used in this study. We would like to thank Aaron Gregoire and the Center for Animal Resources and Education (CARE) staff for their assistance with animal care. We would like to thank John Murphy for his engineering expertise in designing modifications to the sling apparatus. This work was sponsored in part by the Defense Advanced Research Projects Agency (DARPA) BTO under the auspices of Dr. Alfred Emondi and Dr. Jean-Paul Chretien through the (Space and Naval Warfare Systems Center, Pacific OR DARPA Contracts Management Office) Grant/Contract No. D19AC00015 to David Borton. This project was additionally supported by the Brainstorm program within the Carney Institute for Brain Science and the Center for Computation and Visualization (CCV) at Brown University. Computing hardware was supported by NIH Office of the Director Grant S10OD025181.

Received: May 18, 2023
Accepted: August 03, 2023
Published: October 28, 2023

REFERENCES

1. Morris, Richard GM, et al. "Place navigation impaired in rats with hippocampal lesions." *Nature*, vol. 297 18 Jan. 1982, pp. 681–683, https://doi.org/10.1038/297681a0.
2. Mathis, Alexander, et al. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning." *Nat Neuroscience*, vol. 21, 20 Aug. 2018, pp. 1281–1289, https://doi.org/10.1038/s41593-018-0209-y.
3. Nath, Tanmay, et al. "Using DeepLabCut for 3D markerless pose estimation across species and behaviors." *Nat Protocols*, vol. 14, 21 Jun. 2019, pp. 2152–2176, https://doi.org/10.1038/s41596-019-0176-0.
4. Datta, Sandeep Robert et al. "Computational Neuroethology: A Call to Action." *Neuron*, vol. 104, no. 1, 9 Oct. 2019, pp. 11-24. https://doi.org/10.1016/j.neuron.2019.09.038.
5. Hasse, Brady A et al. "Restoration of complex movement in the paralyzed upper limb." *Journal of neural engineering*,

DOI: https://doi.org/10.59720/23-109

vol. 19, no. 4, 1 Jul. 2022, https://doi.org/10.1088/1741-2552/ac7ad7.

6. Labuguen, Rollyn, et al. "MacaquePose: A Novel "In the Wild" Macaque Monkey Pose Dataset for Markerless Motion Capture." *Frontiers in Behavioral Neuroscience*, vol. 14, 18 Jan. 2021, https://doi.org/10.3389/fnbeh.2020.581154.
7. Safayi, Sina, et al. "Kinematic analysis of the gait of adult sheep during treadmill locomotion: Parameter values, allowable total error, and potential for use in evaluating spinal cord injury." *Journal of the Neurological Sciences*, vol. 358, no. 1–2, 2015, pp. 107-112, https://doi.org/10.1016/j.jns.2015.08.031.
8. Karashchuk, Pierre, et al. "Anipose: a toolkit for robust markerless 3D pose estimation." *Cell reports*, vol. 36, no.13, 2021, https://doi.org/10.1016/j.celrep.2021.109730.
9. Graving, Jacob M., et al. "DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning." *eLife*, vol. 8, 1 Oct. 2019, https://doi.org/10.7554/eLife.47994.
10. Becker, A. "Kalman Filtering." (2022), www.kalmanfilter.net/default.aspx.