© 2022. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2022) 225, jeb244540. doi:10.1242/jeb.244540

METHODS & TECHNIQUES

DeepLabCut increases markerless tracking efficiency in X-ray video analysis of rodent locomotion

Nathan J. Kirkpatrick[1,*], Robert J. Butera[1,2] and Young-Hui Chang[1,3]

ABSTRACT

Despite the prevalence of rat models to study human disease and injury, existing methods for quantifying behavior through skeletal movements are problematic owing to skin movement inaccuracies associated with optical video analysis, or require invasive implanted markers or time-consuming manual rotoscoping for X-ray video approaches. We examined the use of a machine learning tool, DeepLabCut, to perform automated, markerless tracking in bi-planar X-ray videos of locomoting rats. Models were trained on 590 pairs of video frames to identify 19 unique skeletal landmarks of the pelvic limb. Accuracy, precision and time savings were assessed. Machine-identified landmarks deviated from manually labeled counterparts by 2.4±0.2 mm (n=1710 landmarks). DeepLabCut decreased analysis time by over three orders of magnitude (1627×) compared with manual labeling. Distribution of these models may enable the processing of a large volume of accurate X-ray kinematics locomotion data in a fraction of the time without requiring surgically implanted markers.

KEY WORDS: Biomechanics, X-ray videography, Rat kinematics, XMA

INTRODUCTION

Despite its utility for relating mechanism to functional outcome, quantifying the movement of behaving rodents can be labor intensive, costly and time consuming. Owing to extensive soft tissue motion artifacts on rodent limbs, optical video tracking approaches using traditional skin-based markers can produce very large joint angle errors up to 39 deg (Bauman and Chang, 2010). In response to this limitation, biplanar high-speed X-ray video has been implemented to directly track the animal skeleton, either via surgically implanted radio-opaque bone-based markers (Brainerd et al., 2010), or by manual alignment (or rotoscoping) of 3D bone models to each pair of video frames from two X-ray camera views (Gatesy et al., 2010). Despite the accurate motion tracking that these X-ray-based methods provide, they each also introduce new limitations of their own. For example, surgically implanting multiple small (sub-millimeter diameter) spheres into each skeletal rigid body is highly invasive and not feasible for some small bones where wound healing may be difficult (e.g. rat foot bones). The surgical wounds created by the marker implantation can introduce a large source of variability in the biomedical research often performed on rats and mice. This added variability can make it difficult to distinguish the effects of the treatment on the injury model from the motion analysis technique. Additionally, rotoscoping 3D bone models typically necessitates expensive µCT scans of each animal that must be manually segmented before beginning the highly labor-intensive process of aligning each bone model by hand.

The recent introduction of a new machine learning tool for pose estimation called DeepLabCut has reduced the analytical burden inherent in markerless tracking for traditional optical video (Mathis et al., 2018). Although others have deployed this new technique to automate the process of tracking surgically implanted markers in X-ray videos (Laurence-Chasen et al., 2020), the ability to identify skeletal landmarks in X-ray video data without markers has yet to be explored. Markerless identification of skeletal landmarks in X-ray video would greatly decrease analysis time while preserving a non-invasive approach and kinematics data quality. This could greatly broaden accessibility to accurate and precise small animal kinematics data.

The goals of this study were to: (1) assess the ability of DeepLabCut to track skeletal landmarks of the rat hindlimb, (2) quantify the time savings compared with manual tracking and (3) develop and distribute the pre-trained DeepLabCut models to track these landmarks in X-ray videos for other researchers to use.

MATERIALS AND METHODS

Model availability

The models described below, as well as their associated labeled training data and processing code, can be accessed under the name ‘xray_rat_hindlimb’ at github.com/njkirkpatrick/xray_rat_hindlimb.

Animal models

Biplanar high speed X-ray videos (Fig. 1A; 45 kV, 100 mA, 100 frames s^{-1}; Imagining Systems & Service, Painesville, OH, USA) were recorded from a total of six adult male Lewis rats [*Rattus norvegicus* (Berkenhout 1769)] performing treadmill locomotion in accordance with protocols approved by the Georgia Institute of Technology’s IACUC. Recordings of a reference object with a known angle moving around the capture volume of our X-ray motion analysis system report a mean error of 0.2 deg (a measure of accuracy) and a variance of 0.8 deg (a measure of precision). Treadmill position within the capture volume varied between recording sessions, but the direction of gait for all trials was oriented towards the left of the frame for both cameras (Figs 1A and 2B).

Data generation

Nineteen hindlimb skeletal landmarks were manually identified in 590 pairs of X-ray video frames taken from 15 videos across the six

[1]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, GA 30332, USA. [2]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. [3]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA.

*Author for correspondence (nk@gatech.edu)

N.J.K., 0000-0002-6263-7387; R.J.B., 0000-0002-1806-0621; Y.-H.C., 0000-0001-7987-6459

Received 11 May 2022; Accepted 23 July 2022

Journal of Experimental Biology (2022) 225, jeb244540. doi:10.1242/jeb.244540

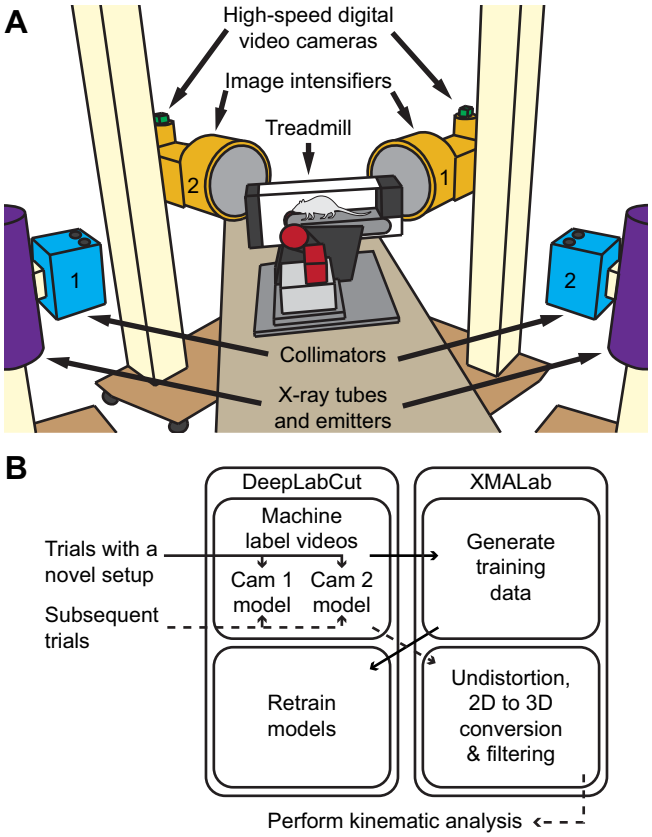


Fig. 1. Schematics of X-ray video collection and skeletal landmark tracking. (A) Configuration of the biplanar X-ray video system. Adapted from Hetzendorfer (2017). (B) Flowchart of the process to train models on a new experimental setup (solid line) and how to analyze trial videos after sufficient training (dashed line). XMALab software is used to correct for fluoroscope distortion as well as to convert and low-pass filter 3D coordinates.

animals (Fig. 2A,B). Training data videos were recorded on eight different days. Owing to naturally occurring differences in treadmill position on each day, collections on different days in this context indicate that there were non-trivial changes in camera perspective from one day to the next. Skeletal landmarks were identified in video pairs within XMALab (Knörlein et al., 2016) to minimize reprojection error between the two views.

The 19 skeletal landmarks were chosen for their visual clarity. The pelvis is tracked with the pubic symphysis. Proximal limb landmarks included the femoral head, greater trochanter, and lateral epicondyle. For the lower limb segment, the lateral condyle of the tibia, distal fusion of the tibia and fibula, and lateral malleolus of the tibia are tracked. Landmarks on the paw are the caudalmost point of the calcaneus bone, distal end of the first metatarsus and distal end of the first phalanx.

Model training

For body part tracking, we used DeepLabCut (version 2.2.0.3) (Mathis et al., 2018; Nath et al., 2019). Labeled videos from each of the two X-ray cameras, as described above, were trained on separate networks. The DeepLabCut default set of 95% of these frames was then used for training. We used a ResNet50-based neural network with default parameters for 314,000 training iterations for each camera model in Google Colab (Mountain View, CA, USA) using our custom fork of XROMMTools (Laurence-Chasen et al., 2020) to convert XMALab data into a DeepLabCut-readable format (see xray_rat_hindlimb GitHub).

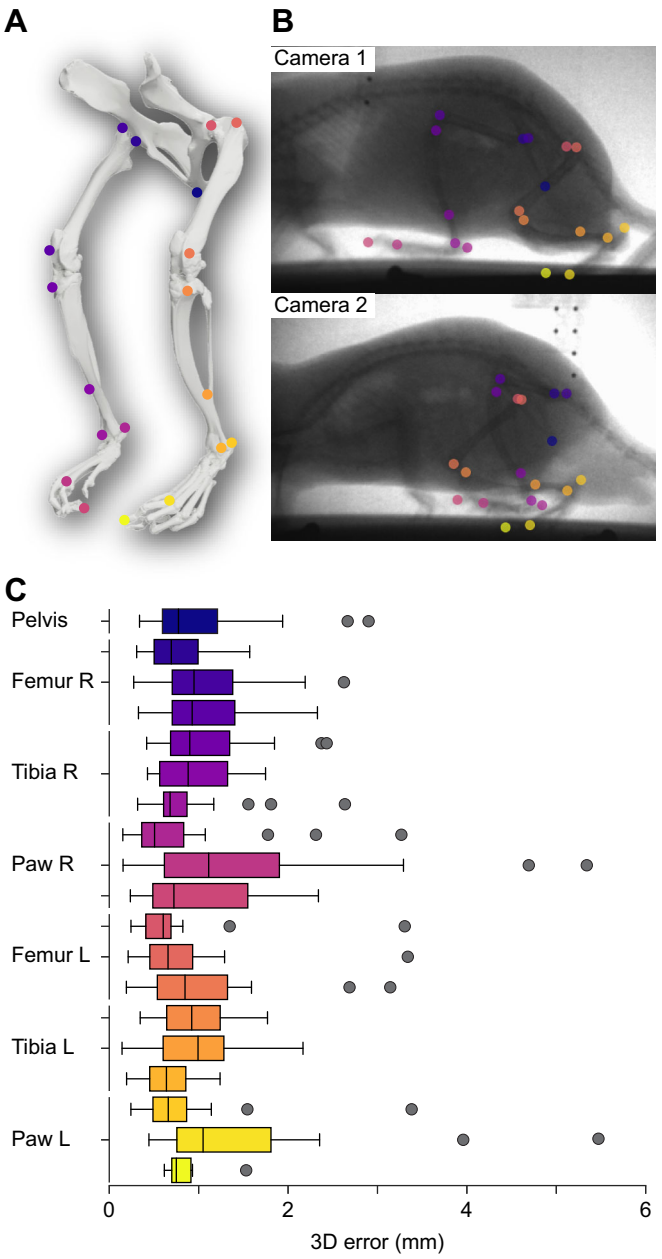


Fig. 2. Model performance relative to manually labeled skeletal landmarks. (A) Locations of rat hindlimb skeletal landmarks used in model training shown in an extended pose for visibility. (B) Sample raw machine labels from a random frame pair; marker colors correspond to those locations indicated in A. (C) Distribution of 3D point error between manually labeled and machine labeled skeletal landmarks of testing dataset frame pairs (n=30 frame pairs). Bars show quartiles with colors corresponding to locations in A. Outliers indicated by circles. 3D points were low-pass filtered at 7 Hz.

We tested our models for consistency by training new models on three separate occasions using a randomized shuffle to generate the 95% training set. For each camera model, we found the test error averaged 5.48±0.70 pixels, and average train error was 2.94±0.11 pixels for the camera 1 network. For the camera 2 network, the test error was 5.56±0.52 pixels and the train error was 3.21±0.25 pixels. Both cameras have image sizes of 1920×900 pixels. We then used the DeepLabCut default p-cutoff of 0.9 to condition the x,y coordinates for future analysis.

Journal of Experimental Biology (2022) 225, jeb244540. doi:10.1242/jeb.244540

This network pair was then used to analyze videos from similar experimental settings.

Undistortion and 3D conversion

XMALab was used to undistort the raw X-ray videos and triangulate machine label pairs of 2D pixel locations into 3D coordinates in lab space. All 3D coordinates were low-pass filtered (7 Hz) to minimize noise using XMALab's built-in low-pass Butterworth filter.

3D point error

For frame pairs included in the testing dataset (n=30 frame pairs), the distance between machine labeled 3D points and their manually labeled counterparts was used as an error metric for the performance of the models (Martin Bland and Altman, 1986). Standard deviation was used throughout.

Joint angle analysis

3D joint angles for the hip, knee and ankle were computed between respective sets of three skeletal landmarks (see Fig. 2A). Hip joint angles were defined by labels from the pubic symphysis, femoral head and lateral epicondyle of the femur. Knee joint angles were defined by labels on the femoral head, lateral epicondyle and lateral malleolus of the tibia. Ankle joint angles consisted of the lateral tibial condyle, lateral malleolus and the distal end of the first metatarsus.

To further evaluate the models, all 19 skeletal landmarks were manually identified in one complete gait cycle (43 frame pairs; Fig. 3A, Movie 1). These manual labels were not included in the model training datasets. The root mean square error was computed between manually labeled and machine labeled joint angles to quantify the discrepancy between the machine labeling and manual labeling techniques. The time required to perform this manual labeling was monitored with a stopwatch and used to estimate time savings.

RESULTS AND DISCUSSION

3D point error

For the 90 frame pairs randomly assigned to the testing datasets, machine label-derived locations in 3D space deviated from manually labeled counterparts by 2.4±0.2 mm (n=1710 skeletal landmarks; Fig. 2C). The 3D point error of the training datasets was 1.9±0.1 mm (n=31,920 skeletal landmarks).

Full gait cycle comparison

Comparing 3D joint angles calculated from machine labels relative to manual labels indicates higher accuracy at the hip and knee joints compared with those of the ankle (Fig. 3A, Movie 1). Over the complete gait cycle, a root mean square error of 1.03 deg was observed for the hip angle and 0.33 deg for the knee, compared with 1.87 deg for the ankle. Machine performance diverged the most for the ankle joint at the end of stance phase near toe off between 60 and 90% of the gait cycle.

Time savings

To demonstrate the type of high-throughput analysis afforded by this method, 83 gait cycles collected on a single day from one rat were analyzed (Fig. 3B). Connected to a GPU in Google Colab, our DeepLabCut models were able to assign 19 labels per frame at a rate of over 11 individual frames (1920×900 pixels) per second (209 landmarks labeled per second). For comparison, a highly trained human took 2 min 24 s to identify those same 19 skeletal landmarks in a single video frame. Extrapolating this average manual analysis

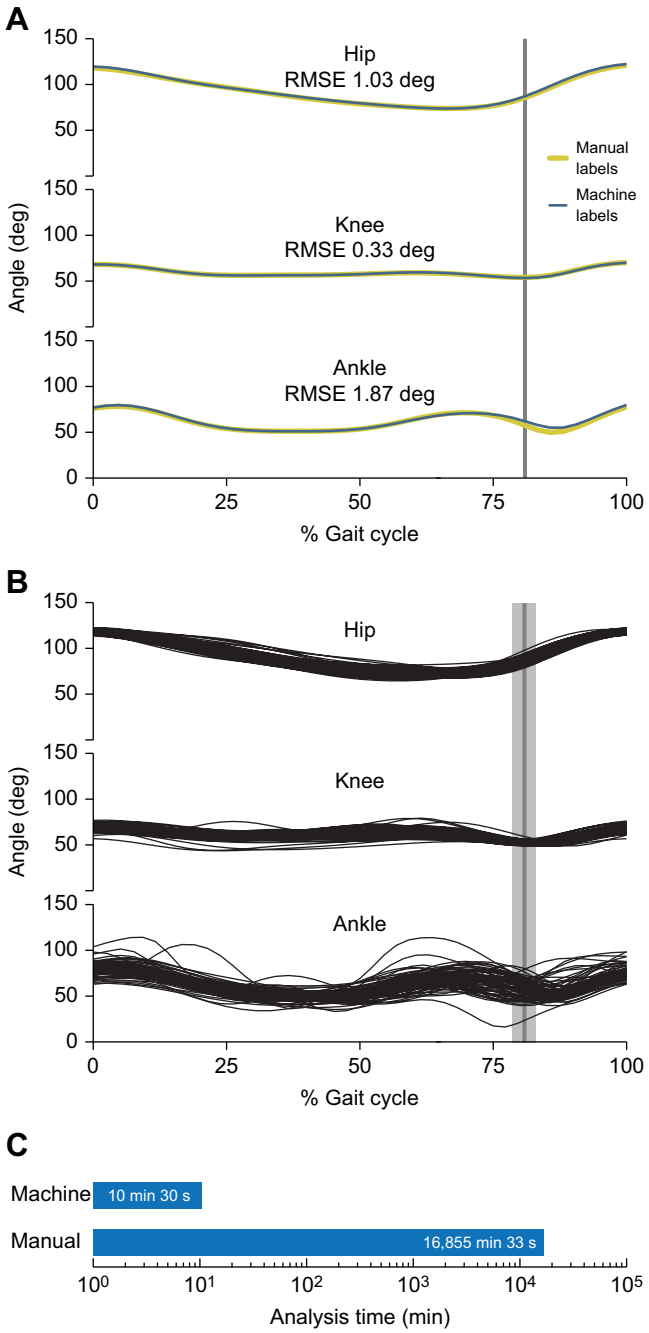


Fig. 3. DeepLabCut allows for robust skeletal tracking without the burden of manual labeling. (A) Representative data showing a direct comparison of left hindlimb sagittal plane joint angles from manually labeled landmarks (blue) and those identified by the models (orange) for one complete gait cycle (n=43 frame pairs, 100 Hz). Root mean square error (RMSE) over the entire gait cycle was computed to quantify the discrepancy between labeling techniques. Vertical line indicates toe-off. A 7 Hz low-pass filter was applied to all data. Labeled frames used for this graph can be seen in Movie 1. (B) Representative model-derived sagittal plane joint angles from 83 gait cycles collected from a single animal during one data collection day. 3D points were low-pass filtered at 7 Hz. (C) Actual processing time for DeepLabCut to label 7014 frames (3507 frame pairs, or 133,266 labels), and the estimated time it would take a trained human to generate the same dataset, indicating a 1627-fold difference in analysis time.

time, it would take a trained human over 11 days, 17 h and 25 min of non-stop work to label the same 7014 frames that DeepLabCut analyzed in 10.5 min (Fig. 3C). In other words, our trained

DeepLabCut models were able to label skeletal landmarks at a rate of 211.5 labels s^{-1}, where a trained human could work at a rate of 0.13 labels s^{-1}, representing a greater than 1600-fold improvement in the rate of analysis.

Process overview

Based on the low variance of testing and training dataset errors between the three models trained on independently shuffled training data, we suggest a single brief retraining of our models until performance plateaus on newly collected experimental data. Training data generation can be expedited by analyzing new trial videos with the our models first, before manually refining the labels in XMALab (Fig. 1B, solid line). Instead of using DeepLabCut's built-in labeling tools, we recommend using XMALab for label refining in order to capitalize on the reprojection error and trajectory graphing features to quickly identify outlier frames that should be manually refined and included in the retraining dataset. While creating the retraining dataset, users should be careful to be consistent in the application of the manual labels, and to provide no coordinates for any labels that are not visible in the given frame. Other previous optical video DeepLabCut projects have successfully performed retraining with only 11 frames to update an existing model to an entirely new viewpoint (case study 2, Mathis et al., 2018). In our experience, DeepLabCut performs substantially better on X-ray video frames when models are trained on each camera independently, rather than a single model for both camera views simultaneously.

After briefly retraining the models, all experimental data can be processed by DeepLabCut to generate machine labels. These labels can then be reformatted using our fork of XROMMTools (Laurence-Chasen et al., 2020) that allows for p-cutoff-based filtering of identified points based on DeepLabCut's likelihood scores. Next, these data are imported into XMALab as distorted 2D coordinates. Once in XMALab, any X-ray distortion can be corrected, and the pairs of 2D pixel coordinates converted to 3D space, then low-pass filtered to reduce jitter and exported for analysis.

Alternatively, some users may want to train new models from scratch using our training data. This may be a good option if, for example, new trial videos include prominently visible forelimbs in the frame and our existing models erroneously label forelimb landmarks. In that case, a pair of new DeepLabCut projects should be created using our fork of XROMMTools (Laurence-Chasen et al., 2020) with forelimb skeletal landmarks included in the list of body parts for DeepLabCut to track. Including locations for forelimb skeletal landmarks in additional training data can reduce the incidence of DeepLabCut misidentifying forelimb skeletal features for desired hindlimb landmarks.

Limitations

Although our DeepLabCut models generate 3D points accurate to 2.4 mm on average (test error), some researchers may require the sub-millimeter accuracy provided by rotoscoping (Gatesy et al., 2010). We posit that the ability to obtain multiple gait cycles from an animal would afford a more accurate representation of average behavior than that of any single gait cycle. Furthermore, the elimination of the additional steps required for rotoscoping may make our machine labeling method an appealing alternative. Additionally, the models may be limited by the quality of the X-ray exposure and ability of the animal to stay in frame, which should be taken into consideration. We suspect that much of the error observed in the labels of the distal limb landmarks in our models was due to occlusion of the paws by our treadmill belt, which may be site-specific.

Use of markerless tracking of skeletal features in X-ray video frames may provide an advantage over implanted markers in the case of machine labeling. In one recent study, DeepLabCut proved to be unsuccessful in tracking surgically implanted, radio-opaque markers in one of the three test cases analyzed (Laurence-Chasen et al., 2020). The failure to track these markers in this study was attributed to the non-cyclical nature of the behavior featured in the examined X-ray videos. Repetitive behavior in trial videos can certainly improve model performance, as a small number of training frames can better represent the variety of the expected movements. However, DeepLabCut has been used to track non-cyclical movements in optical video (Labuguen et al., 2021). Accordingly, the model's insufficient performance in X-ray videos of non-cyclical movements may have been due to the identical appearance of each radio-opaque marker. The difficulty of tracking a novel, non-cyclical behavior may be compounded by the lack of visual distinction between each spherical tantalum marker, leading to a scenario where DeepLabCut cannot sufficiently develop unique classifications for each label. By tracking visually distinct skeletal features instead of uniform spherical markers. DeepLabCut may in fact have a greater likelihood of success in markerless X-ray video.

Comparative advantages

With a three order of magnitude reduction in analysis time and no substantial decrease in accuracy, DeepLabCut using our models can provide the expected quality of X-ray-based kinematics while eliminating a critical rate-limiting step in the current analysis pipeline for markerless X-ray analysis (Fig. 3B). Now, the limiting factor is the amount of video data that can be recorded from a particular animal. By empowering the investigation of more behavioral data from more gait cycles, the strength of a larger dataset could lead to a more accurate view of animal locomotor behavior.

Acknowledgements

We would like to thank Megha Tippur for help with the X-ray system configuration, and Sadie Abernathy for assistance collecting data.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: N.J.K., Y.-H.C.; Methodology: N.J.K.; Software: N.J.K.; Validation: N.J.K.; Formal analysis: N.J.K.; Investigation: N.J.K., Y.-H.C.; Resources: Y.-H.C., R.J.B.; Data curation: N.J.K.; Writing - original draft: N.J.K.; Writing - review & editing: R.J.B., Y.-H.C.; Visualization: N.J.K.; Supervision: R.J.B., Y.-H.C.; Project administration: R.J.B., Y.-H.C.; Funding acquisition: R.J.B.

Funding

This work was not supported by any funding sources external to Georgia Tech.

Data availability

The data and materials for all experiments are available at: www.github.com/njkirkpatrick/xray_rat_hindlimb.

References

Bauman, J. M. and Chang, Y.-H. (2010). High-speed X-ray video demonstrates significant skin movement errors with standard optical kinematics during rat locomotion. *J. Neurosci. Methods* **186**, 18-24. doi:10.1016/j.jneumeth.2009.10.017

Brainerd, E. L., Baier, D. B., Gatesy, S. M., Hedrick, T. L., Metzger, K. A., Gilbert, S. L. and Crisco, J. J. (2010). X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. *J. Exp. Zool. A Ecol. Genet. Physiol.* **313A**, 262-279. doi:10.1002/jez.589

Gatesy, S. M., Baier, D. B., Jenkins, F. A. and Dial, K. P. (2010). Scientific rotoscoping: a morphology-based method of 3-D motion analysis and visualization. *J. Exp. Zool. A Ecol. Genet. Physiol.* **313A**, 244-261. doi:10.1002/jez.588

Hetzendorfer, K. M. (2017). The effects of rehabilitation interventions on hind limb kinematics in a rat model of osteoarthritis. *MSc thesis*, Georgia Institute of Technology, Atlanta, GA.

Knörlein, B. J., Baier, D. B., Gatesy, S. M., Laurence-Chasen, J. D. and Brainerd, E. L. (2016). Validation of XMALab software for marker-based XROMM. *J. Exp. Biol.* **219**, 3701-3711. doi:10.1242/jeb.145383

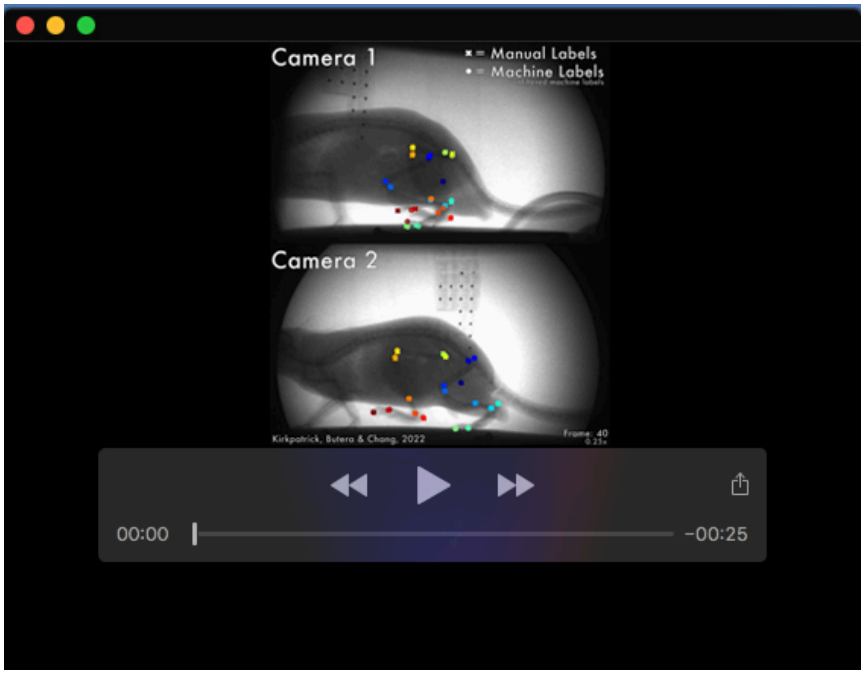
Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., Go, Y., Inoue, K. and Shibata, T. (2021). MacaquePose: a novel 'in the wild' macaque monkey pose dataset for markerless motion capture. *Front. Behav. Neurosci.* **14**, 581154. doi:10.3389/fnbeh.2020.581154

Laurence-Chasen, J. D., Manafzadeh, A. R., Hatsopoulos, N. G., Ross, C. F. and Arce-McShane, F. I. (2020). Integrating XMALab and DeepLabCut for high-throughput XROMM. *J. Exp. Biol.* **223**, jeb226720. doi:10.1242/jeb.226720

Martin Bland, J. and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**, 307-310. doi:10.1016/S0140-6736(86)90837-8

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

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M. and Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **14**, 2152-2176. doi:10.1038/s41596-019-0176-0



Movie 1. Example comparison of manual and machine labeled frames for a complete gait cycle. Unfiltered machine labels (marked with x) compared to manual labels (marked with a dot). Landmarks in this gait cycle were used to compute the joint angles in Fig. 3A. Label colors are consistent with the locations identified in Fig. 2A. X-ray video recorded at 100 fps and played back once at 1.0x speed, then twice at 0.25x for clarity.