Using DeepLabCut to Predict Locations of Subdermal Landmarks from Video

Diya Basrai[1(✉)], Emanuel Andrada[2], Janina Weber[2], Martin S. Fischer[2], and Matthew Tresch[1,3]

[1] Department of Physiology, Northwestern University, Chicago, IL, USA
diyabasrai@gmail.com
[2] Institute of Zoology and Evolutionary Research, Friedrich-Schiller-University, Jena, Germany
[3] Shirley Ryan Ability Lab, Chicago, IL, USA

Abstract. Recent developments in markerless tracking software such as DeepLabCut (DLC) allow estimation of skin landmark positions during behavioral studies. However, studies that require highly accurate skeletal kinematics require estimation of 3D positions of subdermal landmarks such as joint centers of rotation or skeletal features. In many animals, significant slippage between the skin and underlying skeleton makes accurate tracking of skeletal configuration from skin landmarks difficult. While biplanar, high-speed X-ray acquisition cameras offer a way to measure accurate skeletal configuration using tantalum markers and XROMM, this technology is expensive, not widely available, and the manual annotation required is time-consuming. Here, we present an approach that utilizes DLC to estimate subdermal landmarks in a rat from video collected from two standard cameras. By simultaneously recording X-ray and live video of an animal, we train a DLC model to predict the skin locations representing the projected positions of subdermal landmarks obtained from X-ray data. Predicted skin locations from multiple camera views were triangulated to reconstruct depth-accurate positions of subdermal landmarks. We found that DLC was able to estimate skeletal landmarks with good 3D accuracy, suggesting that this might be an approach to provide accurate estimates of skeletal configuration using standard live video.

1 Introduction

DeepLabCut (DLC), an open-source markerless tracking software package, allows automated and robust position estimation of skin landmarks [1], and has been widely adopted for use in extracting kinematics from a wide range of animal behaviors. However, since muscles and tendons act directly on bones and joints, quantifying accurate kinematics often requires estimation of the skeletal structure itself, not the skin above [2].

X-ray reconstruction of moving morphology (XROMM) utilizes high-speed X-ray cameras to obtain accurate 3D estimates of subdermal landmarks, such as skeletal features or joint centers [3]. However, XROMM requires frame-by-frame manual annotation, and subdermal landmarks frequently need to be 'marked' with implanted tantalum beads to show up in X-ray videos. Moreover, because of their expense and complexity, XROMM systems are difficult to use widely across experimental setups.

A. Hunt et al. (Eds.): Living Machines 2022, LNAI 13548, pp. 292–296, 2022.
https://doi.org/10.1007/978-3-031-20470-8_29

Here we present an approach that leverages DLC's ability to estimate skin landmarks to increase the throughput and reduce the labor when quantifying kinematics with XROMM. Instead of training DLC models with hand-annotated frames, we train DLC models with frames that are labeled by projecting the 3D positions of skeletal landmarks obtained with XROMM onto simultaneously obtained frames from live video. Thus, well-trained DLC models can learn to estimate the position of subdermal skeletal features using information present in the live video images. Predicted locations from multiple camera views can then be triangulated to reconstruct depth-accurate positions of subdermal landmarks.

In this paper, we demonstrate preliminary results evaluating this approach on the rat forelimb from an overground running task.

2 Methods

2.1 Data Collection

4 trials of a rat performing an overground running task were recorded by two standard video cameras and two high-speed X-ray acquisition cameras. 'Objective' 3D positions of skeletal features and implanted tantalum landmarks in the rat forelimb were manually annotated, using methods described elsewhere [4].

Training data for the DLC models was generated by projecting the 3D position of subdermal landmarks identified using XROMM onto views from the live camera. Each DLC model, one per camera, was trained with 20 equally spaced frames per trial, from 3 trials total. The fourth trial was held out for use in cross-validation of the method (Fig. 1).

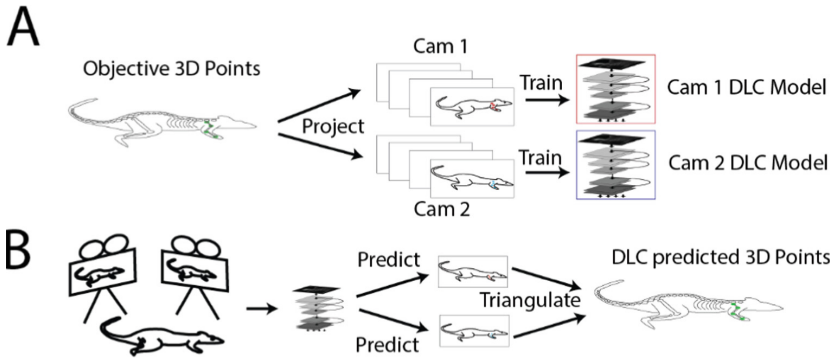

Fig. 1. (A) shows pipeline to generate frames of training data for DLC models. (B) shows pipeline of using trained DLC models to predict 3D positions.

3 Results

3.1 2D Estimation of Skeletal Landmarks

We first evaluated how well DLC models were able to predict the 2D positions of subdermal landmarks. We compared the Euclidian distance between the actual projected points and the DLC predicted points on a trial neither model was trained on. As illustrated in Fig. 2B, we found that errors from this reconstruction were small (~2 pixels).

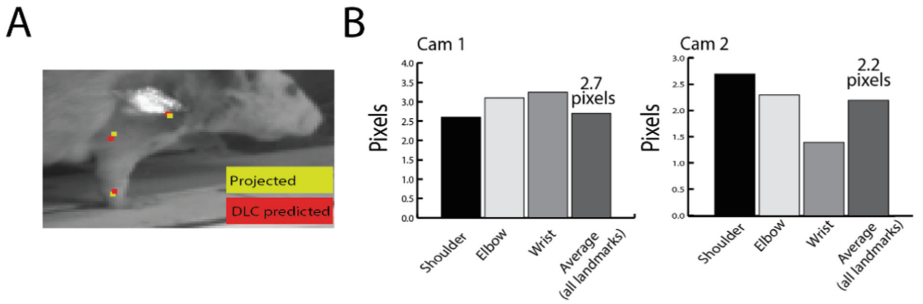

Fig. 2. (A) shows a zoomed in sample frame with projected positions and DLC predicted position. (B) Pixel error for DLC models for certain subdermal landmarks. Pixel error is Euclidean distance between predicted and projected.

3.2 3D Estimation of Skeletal Landmarks

We then triangulated the DLC-predicted points from both cameras to obtain 3D position estimations of subdermal landmarks, and then compared the Euclidian distance between the DLC-predicted 3D points and the objective 3D points obtained by XROMM. As illustrated in Fig. 3A, we found the error to be ~2 mm for each marker.

Finally, we quantified joint angle kinematics from DLC-predicted 3D points and compared those to joint angles obtained from the objective 3D points identified using XROMM. We considered three joint angles and found the average error to be less than ~ 4° (Fig. 3B).

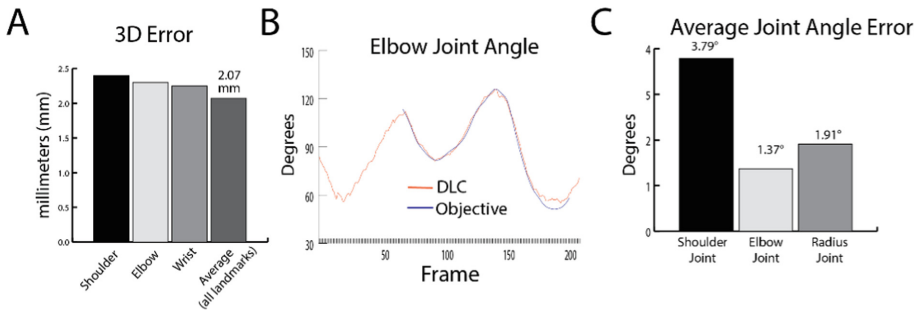

Fig. 3. (A) shows Euclidian distance between DLC triangulated 3D points and objective points. (B) Frame by frame joint angles calculated from DLC predicted landmarks (orange) and from objective (blue) landmarks identified using XROMM. The plot illustrates the joint angle between the shoulder marker, elbow marker, and wrist marker. (C) Average error over all frames for each calculated joint angle. Shoulder joint is calculated between scapular marker, shoulder marker, and wrist marker. Elbow joint is between shoulder marker, elbow marker, and wrist marker. Radius joint is calculated between humerus post marker, radius post landmark and wrist marker (Color figure online).

4 Discussion

Trained DLC models were able to estimate 3D positions of subdermal, skeletal landmarks with low error. At a minimum, this approach can increase the throughput of studies needing 3D position estimates of subdermal landmarks in experiments using XROMM data. DLC models can be trained on data from a small number of frames from each trial in a study, and then used to generate 3D positions with low error. However, for the purpose of reducing labor of analysis, it is debatable whether it might be more efficient to instead apply DLC directly to X-ray data to automate the manual annotation process.

Instead, the promise of the method presented here comes from its potential for generalizability to trials and experiments that do not use XROMM recordings. For instance, in this dataset, the field of view of the live video cameras was wider than the X-ray acquisition cameras, allowing the DLC models to predict frames that the X-ray cameras had no access to. Conceivably, a well-trained DLC model could estimate 3D positions for a specific animal across a wide range of experimental conditions, even tasks where there isn't training data from the X-ray acquisition cameras. Even more speculatively, it might be possible to standardize experimental conditions for a specific species across laboratories closely enough so that multiple investigators could estimate skeletal landmarks with accuracy levels close to those achieved using XROMM but using only standard, off-the-shelf, cameras. In future work, we will evaluate the generalizability of this approach, testing whether it would be possible to create a DLC model for a specific species that can be used across laboratories and experimental conditions to estimate accurate skeletal kinematics.

References

1. Mathis, A., Mamidanna, P., Cury, K.M., et al.: DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21,** 1281–1289 (2018)
2. Bauman, J.M., Chang, Y.H.: High-speed X-ray video demonstrates significant skin movement errors with standard optical kinematics during rat locomotion. J. Neurosci. Methods. **186**(1), 18–24 (2010). https://doi.org/10.1016/j.jneumeth.2009.10.017
3. Brainerd, E.L., et al.: X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. J. Exp. Zool. **313A**, 262–279 (2010)
4. Fischer, M.S., Lehmann, S.V., Andrada, E.: Three-dimensional kinematics of canine hind limbs: in vivo, biplanar, high-frequency fluoroscopic analysis of four breeds during walking and trotting. Sci. Rep. **8**, 16982 (2018). https://doi.org/10.1038/s41598-018-34310-010.1002/jez.589