NEUROSCIENCE

The DANNCE of the rats: a new toolkit for 3D tracking of animal behavior

A new approach tracks animal movements in 3D from multiple camera views using volumetric triangulation, reconciling occlusions and ambiguities present in any one camera view.

Pierre Karashchuk, John C. Tuthill and Bingni W. Brunton

Lured by the scent of a discarded slice, Pizza Rat rouses from her hovel, peeks out into the shadows, and takes a few whiffs. She detects rapidly molding mozzarella and approaches cautiously, whisking the ground. She pauses briefly, using her left back leg to appease the nagging itch on her scabby haunch, then swiftly swings her front right paw to stuff the slice into her jaws. Pizza Rat dances her retreat down the subway stairs until, startled by a filming millennial, she abandons the slice and darts back toward safety. In her 15 seconds of fame (https://youtu.be/UPXUG8q4jKU), Pizza Rat used olfactory, visual and mechanical sensing to coordinate dozens of muscles throughout her body. This sequence of rapid, dexterous movements typifies natural animal behavior. However, the neural control of motor sequences like pizza foraging has been challenging to study, due in large part to an inability to quantify natural behavior in three dimensions (3D).

In recent years, the study of natural animal behaviors has been invigorated by advances in computer vision and artificial neural networks. You may already know and love some of these tools, including the filters that turn you into a talking cat on video chat and image searches that allow you to replace your favorite sandals that were stolen by a seagull on a trip to the beach. Most of these methods work on two-dimensional (2D) images, identifying or tracking objects from a single camera view. But now, a paper in this issue of *Nature Methods* describes a new approach to track animal movement and estimate body pose in 3D[1].

To understand the unique contributions of Dunn, Marshall et al.[1], it is important to place this work in the context of past developments in computer vision. Although it is trivial for most humans to spot a toe or an elbow in a picture, automation of this capability has proceeded at a slow pace for decades. Then, in the last 10 years, the application of artificial neural networks, also known as deep learning, rapidly accelerated the capacity of computers to track human poses in images[2]. Two factors were crucial in this revolution. First, artificial neural networks are built with brain-inspired architectures that learn how to detect poses from raw image data, rather than working from a series of predetermined rules. Second, a few assiduous researchers collected and annotated a large dataset of human images with labeled poses, which allowed the community to focus on algorithms and compare their approaches on a common task. These algorithms were soon harnessed to track movements and estimate body pose from videos of animal behavior[3–5]. But although deep learning has greatly expanded our capacity to measure animal behavior, current tracking methods still have limitations: they typically require manually labeling hundreds of images to fine-tune the neural network, and occlusions remain problematic because they cannot be disambiguated with a single camera view.

Dunn, Marshall et al. address these two challenges by developing a method for 3D animal pose estimation and creating a ground-truth dataset of rat behavior, called Rat 7M. Their DANNCE tracker is among the first to implement an end-to-end algorithm for animal pose estimation in 3D (see also refs. [2,6]). Whereas several recent tools have tackled a similar problem[7–10], they have largely relied on pose tracking in 2D images, followed by synthesis in 3D from multiple cameras using prescribed triangulation rules. In contrast, DANNCE learns pose estimation end to end in 3D from all available camera views by leveraging the shape of the animal, reconciling the uncertainties of poses from any individual camera view (Fig. 1). Furthermore, the Rat 7M dataset consists of 7 million frames of rats behaving in a variety of contexts, all filmed from multiple camera views and labeled with body markers. This dataset will accelerate the training and validation of new algorithms for 3D tracking, particularly for rodents but possibly for other species as well.

With the publication of any new software toolbox, the question naturally

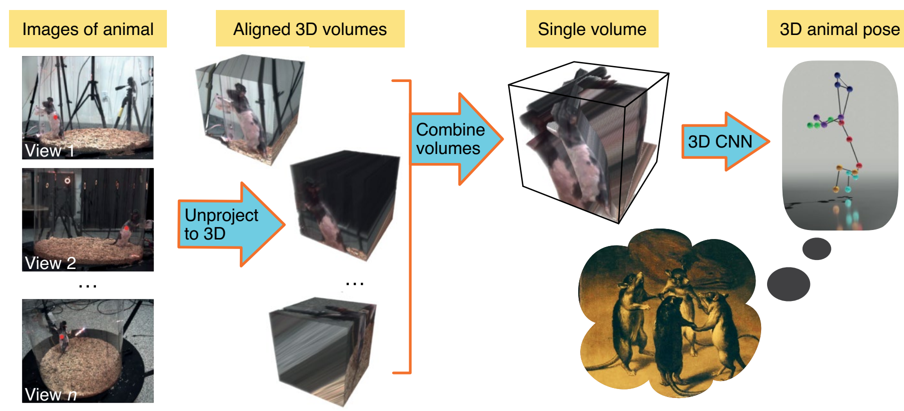

Fig. 1 | DANNCE estimates animal poses in 3D using all available camera views with volumetric triangulation. A single volume serves as input to a 3D convolutional neural network (3D CNN), facilitating reconciliation of occlusions and ambiguities present in any single camera view. Adapted with permission from ref. [1], Springer Nature. Inset: *The Dance of the Rats* (ca. 1690), Ferdinand van Kessel (The Picture Art Collection / Alamy Stock Photo).

arises: what would it take to set it up for my experiments? We believe that there are four key factors that current users of 2D tracking methods should consider when expanding their setup to include 3D tracking with DANNCE. First, it is critical that the cameras are synchronized and calibrated with each other to achieve accurate 3D tracking. Second, a minimal set of images needs to be manually labeled to fine-tune the DANNCE network, and it may be important that these images are similar to those of the animals being studied in size, shape and behavior. Third, the video analysis pipeline needs to be scalable to process many videos—for instance, video of animals in different experimental conditions. Fourth, the tracked positions will still need to be analyzed to gain insights into animal behavior. This step should not be underestimated, and the analysis of high-resolution animal pose dynamics is a rapidly emerging field. Dunn, Marshall et al. contribute initial solutions to each of these hurdles, but individual users may need to clear further technical obstacles specific to their experimental setup, species and behavioral context.

The increasing availability of accessible, robust and high-quality markerless tracking tools is just beginning to transform our understanding of natural behavior in diverse animal species, and many compelling directions are on the horizon. For instance, current approaches struggle with consistency in tracking across consecutive video frames. Not only are smooth trajectories crucial for analyzing body kinematics, temporal information across adjacent frames can be leveraged to improve the quality of pose estimation[9]. Beyond tracking individuals, many animal behaviors are fundamentally social, and some promising tools are already available for 2D tracking of social behaviors from a single camera view[11,12]. Accurate pose estimation of multiple animals interacting in 3D is an exciting frontier.

Building on the availability of large datasets like Rat 7M, the application of semi-supervised and self-supervised learning may drastically reduce the need for manual labels when tracking previously unfamiliar animals. More datasets will be needed to achieve single-camera 3D tracking of other species[13]. In the future, a researcher in the field may be able to take a video of any animal in its natural habitat with a smartphone and extract full 3D kinematics live, a capability that is already becoming available in humans[14]. In addition to improving our understanding of natural animal behavior, real-time tracking technology would enable closed-loop experiments that manipulate neural activity in specific behavioral contexts, including, but not limited to, foraging for scuzzy pizza in the subway. ❒

Pierre Karashchuk[1], John C. Tuthill[2] and Bingni W. Brunton[3]✉
[1]*Neuroscience Graduate Program, University of Washington, Seattle, WA, USA.* [2]*Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA.* [3]*Department of Biology, University of Washington, Seattle, WA, USA.*
✉*e-mail: bbrunton@uw.edu*

Published online: 19 April 2021
https://doi.org/10.1038/s41592-021-01110-w

References
1. Dunn, T.W. et al. *Nat. Methods* https://doi.org/10.1038/s41592-021-01106-6 (2021).
2. Iskakov, K., Burkov, E., Lempitsky, V. & Malkov, Y. Learnable triangulation of human pose. In *Proc. IEEE/CVF Int. Conf. Computer Vision* 7718–7727 (2019).
3. Mathis, A. et al. *Nat. Neurosci.* **21**, 1281–1289 (2018).
4. Pereira, T. D. et al. *Nat. Methods* **16**, 117–125 (2019).
5. Graving, J. M. et al. *eLife* **8**, e47994 (2019).
6. Zimmermann, C., Schneider, A., Alyahyay, M., Brox, T.S. & Diester, I. Preprint at *bioRxiv* https://doi.org/10.1101/2020.02.27.967620 (2020).
7. Nath, T. et al. *Nat. Protoc.* **14**, 2152–2176 (2019).
8. Günel, S. et al. *eLife* **8**, e48571 (2019).
9. Karashchuk, P. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.26.117325 (2020).
10. Bala, P. C. et al. *Nat. Commun.* **11**, 1–12 (2020).
11. Pereira, T. D. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.31.276246 (2020).
12. Chen, Z. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.04.405159 (2020).
13. Gosztolai, A. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.18.292680 (2020).
14. Mehta, D. et al. *ACM Trans. Graph.* **36**, 1–14 (2017).

Competing interests
The authors declare no competing interests.