Deep Graph Pose: a semi-supervised deep graphical model for improved animal pose tracking

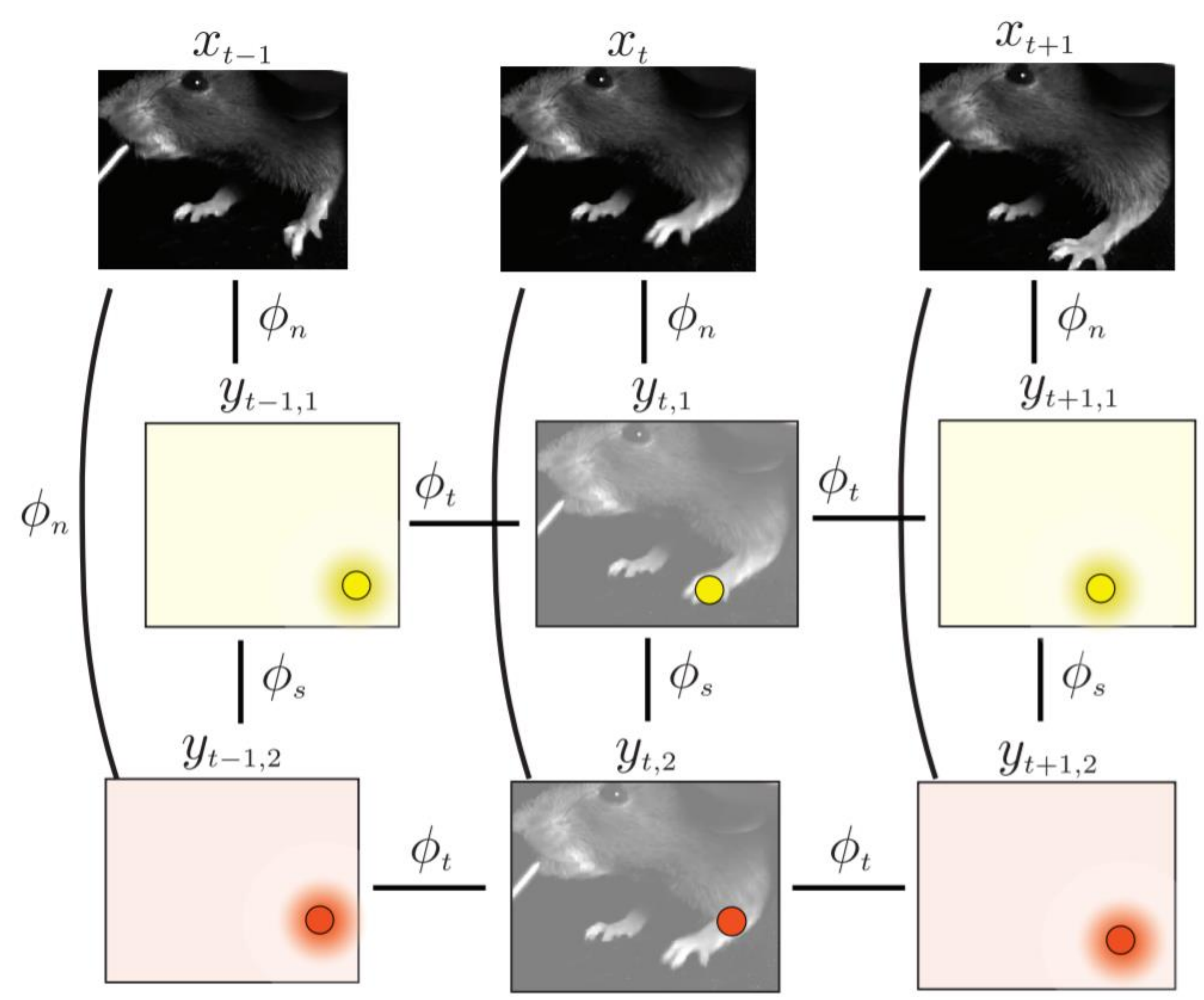
Daiki Tagami[1], Anqi Wu[1], Kelly Buchanan[1], Dan Biderman[1], Sunand Raghupathi[1], Matthew Whiteway[1] and Liam Paninski[1]

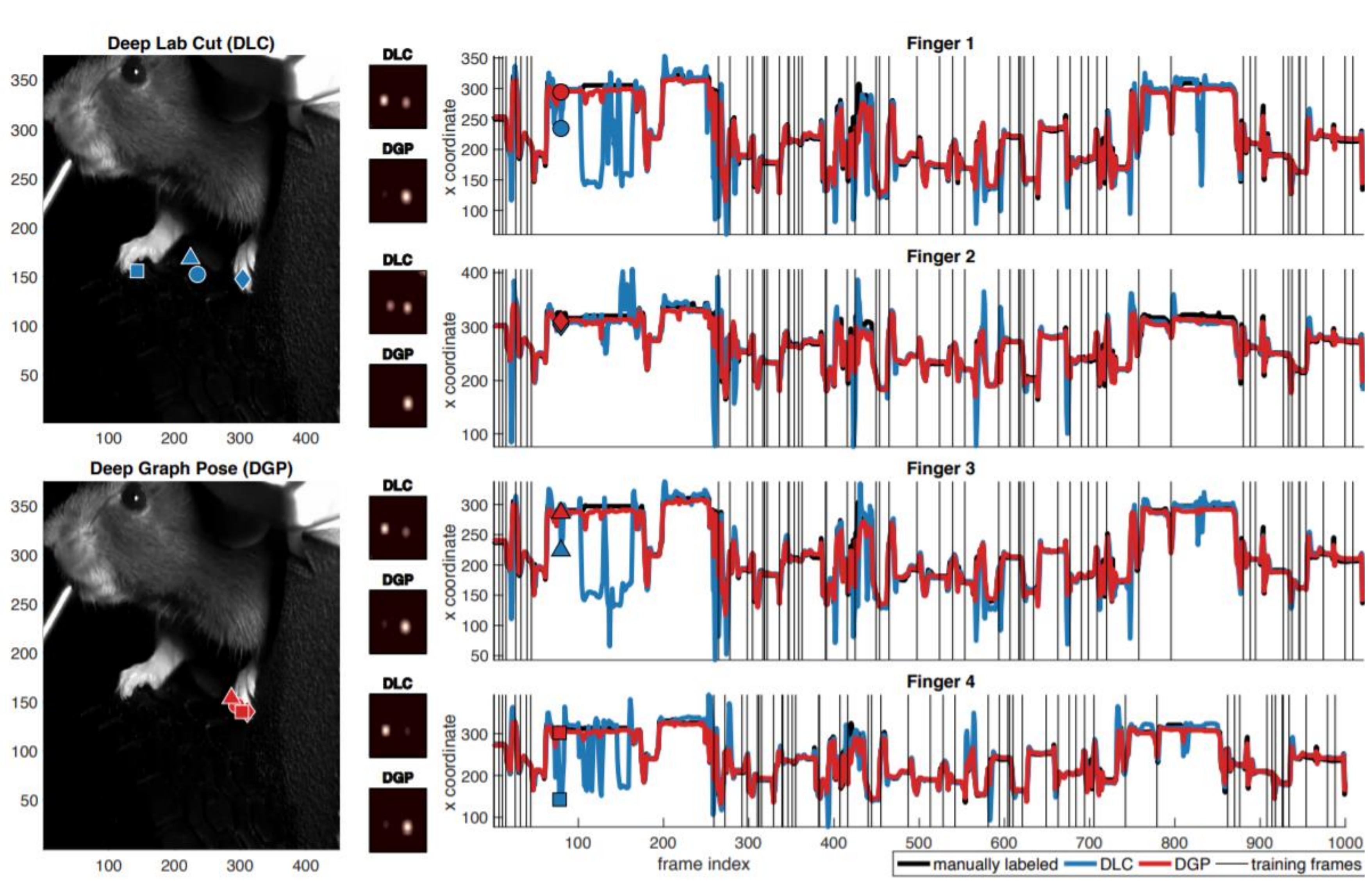
[1]Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, USA

Abstract

Animal pose tracking is crucial for many scientific investigations, with applications in ethology, psychology, neuroscience, and other fields. State of the art methods such as Deep Lab Cut and DeepPoseKit have opened up an exciting array of new applications in the area of animal behavior tracking, but hundreds of labels may still be needed to achieve tracking at the desired level of precision and reliability. Here, we propose Deep Graph Pose, which is a probabilistic graphical model built on top of deep neural networks, and develop an efficient structured variational approach to perform inference with this model. Deep Graph Pose models the targets as continuous random variables, resulting in a semi-supervised model that make use of both labeled and unlabeled frames of animals to achieve significantly higher accuracy while requiring fewer labeled training frames.

Many animal tracking tasks are taking frames from multiple views, and there are many occlusions and missing body parts per frame. While some existing algorithms are doing post-hoc 3D reconstruction, there still exists some issues with creating a model that can completely perform the tracking process in 3D. Here, we try to integrate both the multi view constraint and 3D body reconstruction into Deep Graph Pose, so that the entire tracking model will be 3D.





Comparisons with conventional models

When Deep Graph Pose was compared with Deep Lab Cut, which is one of the state of art animal pose estimation methods, Deep Graph Pose performed better compared to Deep Lab Cut. The vertical lines indicate the labeled frames that were used for training.

Probability Distribution

The joint probability distribution over target y is given as:

$$p(y|x,\beta) = \frac{1}{Z(x,\beta)} \exp\left(-\sum_{t=1}^{T}\sum_{j=1}^{J} \phi_n^j(y_{t,j}, x_t) - \sum_{t=1}^{T-1}\sum_{j=1}^{J} \phi_t^j(y_{t,j}, y_{t+1,j}) - \sum_{t=1}^{T}\sum_{i,j\in\mathcal{E}} \phi_s^{ij}(y_{t,i}, y_{t,j}) \right)$$

Here, $\beta = \{\theta, w_t, w_s\}$ is the parameter vector, where θ denotes the neural net parameters in ϕ_n. ϕ_n is the potential that is defined by the neural network, and a simple quadratic potential was used to encode these soft constraints:

$$\phi_s^{ij}(y_{t,i}, y_{t,j}) = \frac{1}{2} w_s^{ij} ||y_{t,i} - y_{t,j}||^2$$

Acknowledgements

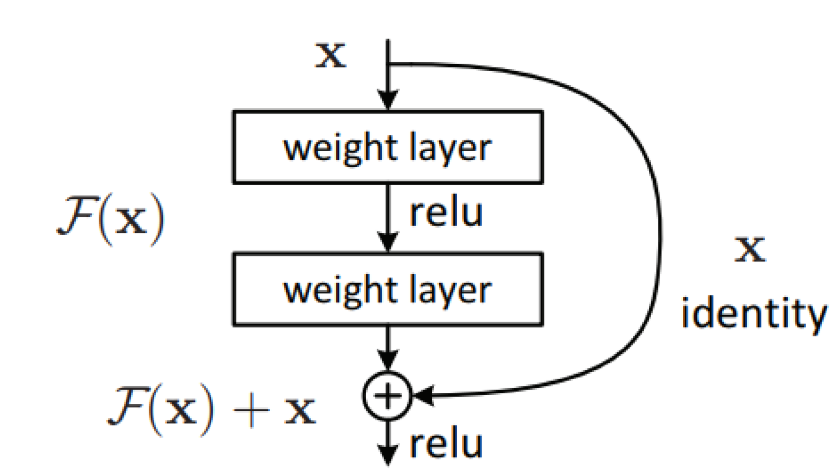
I thank Columbia University in the City of New York, Rabi Scholars Program, and Laidlaw Scholars Program for their generous support. I would also like to thank Dan Biderman and Liam Paninski and all the lab mambers for giving me the opportunity to work on this project. We also thank High Performance Computing (HPC) resources from Columbia University.

Deep Graph Pose Model

Deep Graph Pose uses graph semisupervised inference to estimate the locations of unobserved targets from the observed (labeled) data. Here, x_t indicates the frame that we observe, and we would like to track multiple targets in each frame. $y_{t,1}$ and $y_{t,2}$ shows the observed labels of the two targets (paw and elbow) in some frames. The hidden variables here are the unobserved targets, which are indicated with colored circles in the colored background.

Deep Learning Model

Pre-trained ResNet50 model and trainable convolutional neural network was used to track animal behavior. ResNet is a deep learning model that is frequently used in image recognition purposes, and the structure is shown below:



The probability distribution can be rewritten as a product of neural network and Gaussian graphical model:

$$p(y|x,\beta) = \frac{1}{Z(x,\beta)} \exp\left(\underbrace{-\sum_{t=1}^{T}\sum_{j=1}^{J} \phi_n^j(y_{t,j}, x_t)}_{\text{neural network}} \underbrace{- \sum_{t=1}^{T-1}\sum_{j=1}^{J} \phi_t^j(y_{t,j}, y_{t+1,j}) - \sum_{t=1}^{T}\sum_{i,j\in\mathcal{E}} \phi_s^{ij}(y_{t,i}, y_{t,j})}_{\text{Gaussian graphical model}} \right)$$

References

- Wu, A., Buchanan, E. K., Whiteway, M., Schartner, M., Meijer, G., Noel, J. P., & Paninski, L. (2020). Deep Graph Pose: a semi-supervised deep graphical model for improved animal pose tracking. bioRxiv.
- 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. Nature neuroscience, 21(9), 1281-1289.