The Journal of Open Source Software

TrackSegNet: a tool for trajectory segmentation into diffusive states using supervised deep learning

Hélène Kabbech [1] **and Ihor Smal** [1]

1 Department of Cell Biology, Erasmus University Medical Center, Rotterdam, the Netherlands

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Kevin M. Moerman

Reviewers:

- @imagejan
- @ajasja

Submitted: 03 May 2023
Published: unpublished

Summary

TrackSegNet is a command-line python program, which permits the classification and segmentation of trajectories into diffusive states. A deep neural network is trained for each particular case using synthetic data and trajectory features as inputs. After classification on the experimental data using the trained network, the trajectories are segmentated and grouped per diffusive state. TrackSegNet further estimates the motion parameters (the diffusion constant D and anomalous exponent α) for each segmented track using the mean squared displacement (MSD) analysis, and computes additional geometric measurements per tracklet state such as the angular distribution and velocity autocorrelation curve. The resulting segmentation and motion parameters are stored as CSV files. Originally developped for the quantification of protein dynamics using single-particle tracking imaging, its use can be extended to any type of trajectory dataset.

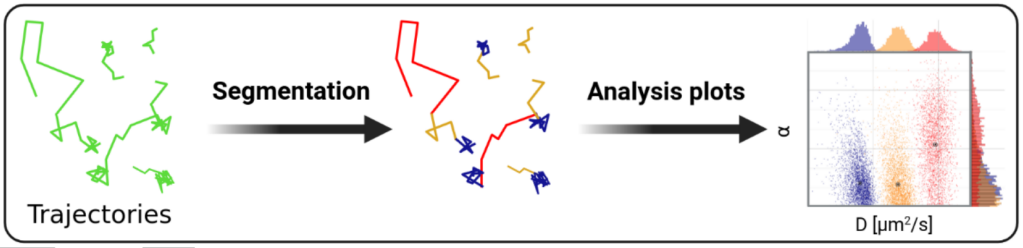

Figure 1: Analysis pipeline of TrackSegNet described in two steps.

Statement of need

Recent advances in the field of microscopy allow the capture, at nanometer resolution, of the motion of fluorescently-labeled particles in live cells such as proteins or chromatin loci. Therefore, the development of methods to characterize the dynamics of a group of particles has become more than necessary (Muñoz-Gil et al., 2021). A typical analysis is the classification and segmentation of trajectories into diverse diffusive states when multiple types of motion are present in a dataset (e.g., confined, superdiffusive) due to the properties of the labeled molecule (e.g., protein bound/unbound to the DNA). A few trajectory classification methods have recently been developed by the community, among which can be cited Wagner et al. (2017), Hansen et al. (2018), Arts et al. (2019), Pinholt et al. (2021) and Kabbech & Smal (2022). However, more practical analysis software is needed for direct application use.

The Journal of Open Source Software

Method

This software is based on the method of Arts et al. (2019) with major improvements and allows replicability on other datasets.

Neural Network

Tracking particles from 2-dimensional images results in a set $\mathcal{S}$ of trajectories $r_i \in \mathcal{S}$, $i = \{1, \dots, P\}$, where P is the total number of trajectories and $r_i(t) = (x_i(t), y_i(t))$ are the 2D coordinates of the particle i at time t.

The network is built using functions from the Keras library (Chollet & others, 2015), and is composed of a bidirectional long short-term memory (LSTM) layer (having 200 hidden units) followed by a fully connected time-distributed layer with a `SoftMax` activation function. The inputs of the network are of six trajectory features computed beforehand, while the outputs are probabilities for each trajectory point of belonging to one of the N diffusive states, the predicted state is defined by the highest probability.

The computed features along a given trajectory are: the displacements $\Delta x_{\delta=1}$ and $\Delta y_{\delta=1}$ at the first discrete time interval $\delta = 1$ (with $\Delta r_\delta(t) = r(t) - r(t+\delta)$), the distances $d_{\delta=1}$ (with $d_\delta(t) = \sqrt{\Delta x_\delta(t)^2 + \Delta y_\delta(t)^2}$), the mean of displacements $\overline{d_{\delta=1,p=1}}$ and $\overline{d_{\delta=2,p=1}}$ (with $\overline{d_{\delta,p}}(t) = \frac{1}{2p+1}\sum_{k=t-p}^{t+p} d_\delta(k)$ with $p \geq 1$) and the angles $\theta_{\delta=1}$ between two consecutive displacements. The last feature is an addition to the initial method, used for a better distinction of the trajectory confinement. The first and last trajectory points of each trajectory vector are discarded due to missing computed feature(s).

Training

The network is trained using synthetic fractional Brownian motion (fBm) trajectories of mixed diffusive states. For this purpose, 10,000 fBm trajectories with a switching mode between states and a total length of 27 frames are generated for each independent training. The fBm process is characterized using the fBm kernel (Lundahl et al., 1986) defined as $k_{\mathrm{FBM}}(t) = D\,[\,|t+1|^\alpha - 2|t|^\alpha + |t-1|^\alpha]$, with $t = \Delta t/\delta$ (Δt the time measured between two frames) and the pre-defined motion parameters $m = (D, \alpha)$.

The model is optimized using `Adam` during the training and a categorical cross-entropy loss function.

Model parameters

The main parameters of the training are tunable from the `params.csv` file to create a new variant of the model:

- `num_states` is an important parameter permitting to decide the number N of diffusive states. This number can vary from 2 to 6 states, but it is recommended to choose 2 to 4 states.
- `state_i_diff` and `state_i_alpha` the approximate motion parameters m for each of the N diffusive state. The diffusion constant D is dimensionless, and the anomalous exponent value α is ranging from 0 to 2 (]0-1[: subdiffusion, 1: Brownian motion,]1-2[: superdiffusion).
- `pt_i_j` the probability of transitionning from the state i to the state j. The total number of probabilities should be N^2.

The remaining parameters are related to the experimental dataset:

- `data_path`, the path of the dataset of trajectories to segment.
- `track_format`, the format of the files containing the trajectory coordinates, either `MDF` (see `MTrackJ` data file format) or `CSV`

- `time_frame`, the time interval between two trajectory points in seconds.
- `pixel_size`, the dimension of a pixel in μm.

Classification and MSD analysis

Before computing the features for each experimental trajectory, gaps in trajectories of length 1 are filled by a randomly generated point; while the larger gaps are split in two separate trajectories. Each point is therefore classified as one of the N diffusive states using the trained LSTM model. Based on the state classification, the trajectories are segmented and the motion parameters are estimated for each segmented track (longer than 5 frames) using the MSD analysis. The latter consists of applying a least-square fit from the logarithm form of the MSD power-law equation (Metzler et al., 2014). Both D and α distributions can be plotted in a scatterplot as shown in Figure 1. The new probability transition matrix and proportion of tracklet points in each diffusive state are also calculated.

Acknowledgements

We would like to thanks Selçuk Yavuz and Martin E. van Royen for sharing SPT data used as a toy example, and Maarten W. Paul for testing the software and fixing minor mistakes. Figure 1 was partially created using Biorender. This work was supported by the Dutch Research Council (NWO) through the Building Blocks of Life research program (GENOMETRACK project, Grant No. 737.016.014).

References

Arts, M., Smal, I., Paul, M. W., Wyman, C., & Meijering, E. (2019). Particle mobility analysis using deep learning and the moment scaling spectrum. *Scientific Reports*, *9*(1), 17160. https://doi.org/10.1038/s41598-019-53663-8

Chollet, F., & others. (2015). *Keras*. https://keras.io.

Hansen, A. S., Woringer, M., Grimm, J. B., Lavis, L. D., Tjian, R., & Darzacq, X. (2018). Robust model-based analysis of single-particle tracking experiments with spot-on. *Elife*, *7*, e33125. https://doi.org/10.7554/eLife.33125

Kabbech, H., & Smal, I. (2022). Identification of diffusive states in tracking applications using unsupervised deep learning methods. *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, 1–4. https://doi.org/10.1109/ISBI52829.2022.9761672

Lundahl, T., Ohley, W. J., Kay, S. M., & Siffert, R. (1986). Fractional brownian motion: A maximum likelihood estimator and its application to image texture. *IEEE Transactions on Medical Imaging*, *5*(3), 152–161. https://doi.org/10.1109/TMI.1986.4307764

Metzler, R., Jeon, J. H., Cherstvy, A. G., & Barkai, E. (2014). Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, *16*(44), 24128–24164. https://doi.org/10.1039/C4CP03465A

Muñoz-Gil, G., Volpe, G., Garcia-March, M. A., Aghion, E., Argun, A., Hong, C. B., Bland, T., Bo, S., Conejero, J. A., Firbas, N., & others. (2021). Objective comparison of methods to decode anomalous diffusion. *Nature Communications*, *12*(1), 6253. https://doi.org/10.1038/s41467-021-26320-w

Pinholt, H. D., Bohr, S. S. R., Iversen, J. F., Boomsma, W., & Hatzakis, N. S. (2021). Single-particle diffusional fingerprinting: A machine-learning framework for quantitative analysis of heterogeneous diffusion. *Proceedings of the National Academy of Sciences*, *118*(31), e2104624118. https://doi.org/10.1073/pnas.2104624118

Wagner, T., Kroll, A., Haramagatti, C. R., Lipinski, H. G., & Wiemann, M. (2017). Classification and segmentation of nanoparticle diffusion trajectories in cellular micro environments. *PloS One*, *12*(1), e0170165. https://doi.org/10.1371/journal.pone.0170165