Supplementary: Learning nanoscale motion patterns of vesicles in living cells

Arif Ahmed Sekh[1] Ida Sundvor Opstad[1] Åsa Birna Birgisdottir[1,2] Truls Myrmel[1,2]
Balpreet Singh Ahluwalia[1] Krishna Agarwal[1] Dilip K. Prasad[1*]
[1]UiT The Arctic University of Norway, Tromsø, Norway
[2] University Hospital of North Norway, Tromsø, Norway
[*]dilip.prasad@uit.no

1. Description of supplementary video

One single supplementary video is provided, which is a stitched slide presentation. The slides are numbered as #1 to #4 and are described below.

#1 presents the steps of physics based simulation of microscopy videos of vesicles and their motion patterns. This illustrates the steps shown in Fig. 3 of the main manuscript.

#2 presents a sample of an experimental live-cell sequence, a sample ROI extracted from the live-cell sequence, and the nanoscopy image of the ROI reconstructed using MUSICAL.

#3: presents a comparison of the feature representations of vesicle motion. The raw image sequence for a vesicle in a live-cell video, the corresponding micro-motion magnified sequence, and the nanoscopy image reconstructed by MUSICAL illustrate that the nanoscopy image provides better feature of identifying the motion pattern.

#4: presents the example shown in Fig. 12 of the main manuscript. The original video of a vesicle moving and interacting with a mitochondrion is shown along with the nanoscopy image obtained by stitching the MUSICAL results of the different sub ROIs.

2. SRAN architecture

The architecture of SRAN is described in Table 1. We use two types of activation function for mixed attention and spatial attention. We have not used three channel attention as reported in [10] because nanoscopy images are single channel. The mixed attention (f_1) is a simple sigmoid over all spatial position. The spatial attention (f_2) is normalized over each channel and then a sigmoid to get the soft mask, where μ is the mean and σ is the standard deviation.

$$f_1(x_i, c) = \frac{1}{1 + \exp(-x_{i,c})} \tag{1}$$

$$f_2(x_i, c) = \frac{1}{1 + \exp(-(x_{i,c} - \mu_c)/\sigma_c)} \tag{2}$$

Table 1. Shallow Residual Attention Network (SRAN) architecture details for nanoscale motion classification. The attention module is built by pre-activation residual units [5] with a number of channel in each stage same as ResNet [4].

Layer	Output Size	Shallow Attention
Input (64×64)	-	-
Conv1	32×32	5×5, 32, stride 2
Max pooling	16×16	2×2, stride 2
Residual Unit	16×16	$\begin{pmatrix} 1 \times 1, 32 \\ 2 \times 2, 32 \\ 1 \times 1, 128 \end{pmatrix}$
Attention Module	16×16	Attention $\times 1$
Residual Unit	8×8	$\begin{pmatrix} 1 \times 1, 64 \\ 1 \times 1, 64 \\ 1 \times 1, 256 \end{pmatrix}$
Attention Module	8×8	Attention $\times 1$
Average pooling	1×1	8×8, stride 1
FC,Softmax	1000, 5	-

3. Baseline hyperparameters for classification methods

Here we present the hyperparameters used for baseline classification. These parameters apply to all the baseline methods used in the paper. We have used standard data augmentation, which include shift, flip, rotation, and brightness changes. We have taken a batch size 32 and adopted early stopping if validation accuracy does not improve for 100 iterations. The learning rate is decreased by a factor of 10 when the validation accuracy stops improving for 20 iterations. The network architecture is adjusted only to fit the input images, and the number of hidden layers and other internal parameters remain the same as the original implementations. Parameters other than the network architecture is presented in Table 2.

Table 2. Hyperparameters of different baseline classification models

Method	Hyperparameters
Deep CNN [6]	Dropout 0.5 in first two fully connected layers. Learning rate 0.01. Weight decay 0.0005. Momentum 0.9. ReLU activation.
VGG16 [13]	Dropout: 0.5 in first two fully connected layers. Learning rate 0.01. Weight decay 0.0005 (L_2). Momentum 0.9. ReLU activation.
Attention Model [12]	Dropout: 0.5. Stochastic gradient descent (SGD). Learning rate 0.01. Weight decay 0.0005. Momentum 0.9. Hypertangent activation.
Shallow Network [7]	Number of hidden units (M)=40000. Filter size (W)=7. Pooling size (Q)=7. Downsample factor (D)=2. Learning rate 0.01.
ResNet50 [4]	Dropout 0.5. SGD. Learning rate 0.01. Weight decay 0.0005. Momentum 0.9.
ResNet20 [4]	SGD. Learning rate 0.01. Weight decay 0.0005. Momentum 0.9.
MLP (Baysian Optimization) [9]	Final MLP contains 3 layers. Each layers has 512 hidden unit. Dropout 0.5 for first two layers. ReLU activation.
Inception V3 [11]	Learning rate 0.01. Weight decay 0.0005. Momentum 0.9.
Deep residual attention [10]	Dropout 0.6. SGD. Learning rate 0.00078. Weight decay 0.0001. Momentum 0.9.
Proposed SRAN	Dropout 0.6. SGD. Learning rate to 0.00047. Weight decay 0.0001. Momentum 0.9.

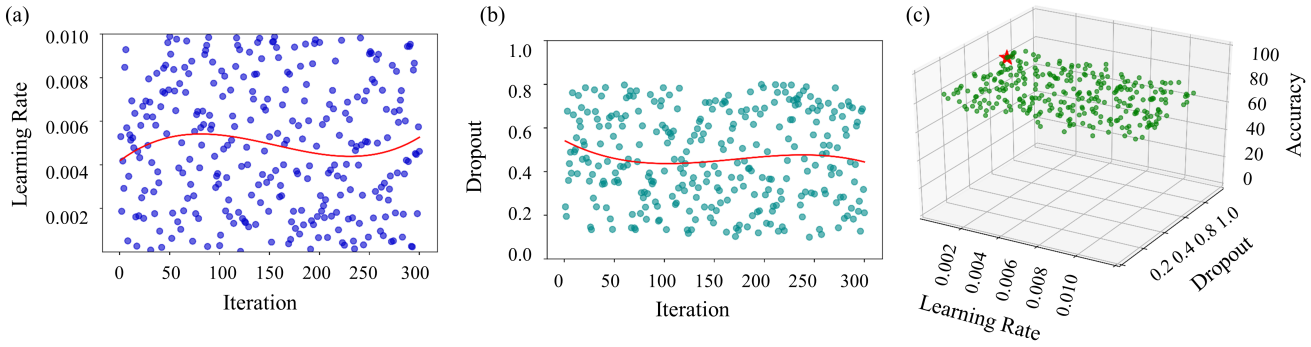

Figure 1. Baysian optimization in deep residual network (a) distribution of learning rate over iteration (b) distribution of dropout over iteration (c) learning rate, dropout and accuracy, red star represents maximum accuracy.

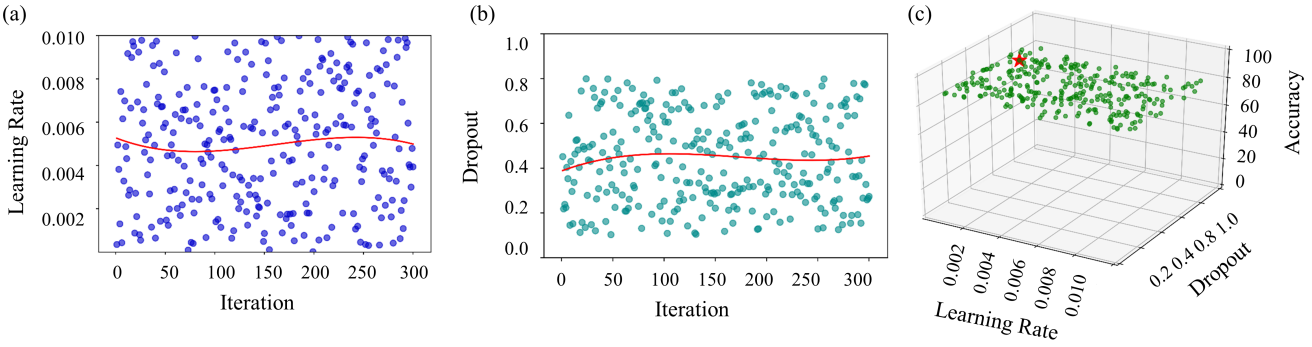

Figure 2. Baysian optimization in proposed shallow residual network (a) distribution of learning rate over iteration (b) distribution of dropout over iteration (c) learning rate, dropout and accuracy, red star represents maximum accuracy.

4. Hyperparameter optimization experiment for DRAN and SRAN

We used Bayesian optimization to fine tune the minimum learning rate and dropout before the dense layer for DRAN and SRAN. The minimum learning rate is in the range $[1e-7, 1e-2]$, and the dropout is in the rang of $[0.1, 0.8]$. We have used 300 iterations for optimization in the simulated data with 70% training and 30% validation setup. Figure 1 shows the evolution of the learning rate and dropout with the iterations and the accuracy as a function of learning rate and dropout for DRAN. Figure 2 present analogous plots for SRAN. It is observed that DRAN performs best with an accuracy of 0.85 with the initial learning rate 0.00078 and dropout 0.6. The shallow network produces 0.89 accuracy with 0.00047 initial learning rate and 0.5 dropout. We also

Table 3. Values of mean and standard deviation (in bracket) of the motion states observed in our live cell dataset. Rows are initial states and columns are subsequent states. Colors match the colors of the box plots in Fig. 11 (a) of the main paper.

	CirC	RCir	Flow	Ranw	Stat
Normal	0.181 (0.039)	0.068 (0.006)	0.063 (0.033)	0.603 (0.086)	0.083 (0.012)
Hypoxia	0.162 (0.017)	0.060 (0.024)	0.045 (0.017)	0.679 (0.021)	0.052 (0.009)
HypoxiaADM	0.110 (0.043)	0.049 (0.009)	0.030 (0.028)	0.739 (0.021)	0.070 (0.015)

Table 4. Values of mean and standard deviation (in bracket) of consecutive-motion-state-pairs observed in our live cell dataset. Rows are initial states and columns are subsequent states. Colors match the colors of the box plots in Fig. 11 (b) of the main paper.

	Initial State	Subsequent State				
		CirC	**RCir**	**Flow**	**RanW**	**Stat**
Normal	**CirC**	0.092 (0.017)	0.019 (0.007)	0.007 (0.005)	0.071 (0.007)	0.009 (0.006)
	RCir	0.016 (0.008)	0.018 (0.005)	0.003 (0.001)	0.037 (0.010)	0.003 (0.003)
	Flow	0.010 (0.010)	0.003 (0.001)	0.011 (0.008)	0.034 (0.020)	0.004 (0.005)
	RanW	0.070 (0.011)	0.034 (0.009)	0.026 (0.017)	0.409 (0.120)	0.037 (0.007)
	Stat	0.008 (0.002)	0.005 (0.001)	0.004 (0.003)	0.034 (0.007)	0.024 (0.004)
Hypoxia	**CirC**	0.076 (0.011)	0.020 (0.008)	0.003 (0.002)	0.083 (0.009)	0.004 (0.002)
	RCir	0.018 (0.013)	0.017 (0.013)	0.001 (0.001)	0.033 (0.004)	0.002 (0.001)
	Flow	0.003 (0.003)	0.001 (0.001)	0.008 (0.007)	0.031 (0.007)	0.001 (0.002)
	RanW	0.075 (0.003)	0.030 (0.008)	0.023 (0.009)	0.480 (0.019)	0.027 (0.015)
	Stat	0.005 (0.002)	0.001 (0.001)	0.001 (0.001)	0.030 (0.004)	0.015 (0.002)
HypoxiaADM	**CirC**	0.048 (0.023)	0.009 (0.004)	0.001 (0.001)	0.060 (0.027)	0.007 (0.002)
	RCir	0.008 (0.004)	0.014 (0.007)	0.001 (0.002)	0.031 (0.002)	0.004 (0.003)
	Flow	0.000 (0.000)	0.002 (0.002)	0.006 (0.000)	0.018 (0.014)	0.002 (0.002)
	RanW	0.054 (0.016)	0.025 (0.003)	0.021 (0.018)	0.579 (0.041)	0.036 (0.009)
	Stat	0.002 (0.002)	0.004 (0.001)	0.001 (0.002)	0.042 (0.009)	0.016 (0.005)

note that the methods are not too sensitive to the learning rate and dropout as the standard deviation of accuracy is 3.62, 2.38 for DRAN and SRAN, respectively.

5. Detailed statistics of state changes

The statistics of the frequency of occurrences of motion states and consecutive-motion-state-pairs were presented graphically in the Fig. 11 of the main manuscript through box plots. Here, we present the values of the mean and standard deviation of each box plot. The statistics for the motion states are presented in Table 3. The statistics for the consecutive-motion-state-pairs are presented in Table 4.

6. Live cell data collection

6.1. Sample preparation

The abbreviations used in this subsection are related to chemical names of associated chemicals used as they are in current scientific literature. Therefore, their complete forms are not included. The rat cardiomyoblast cell-line H9c2 (cells derived from embryonic heart tissue; Sigma Aldrich) was cultured in DMEM with 10% fetal bovine serum. The cells were transiently transfected using TransIT –LT1 (Mirus) to express the mitochondrial fluorescence marker mCherry-OMP25-TM. After 24 hours of transfection, the cells were incubated in serum free DMEM medium for 4 hours and then the medium was changed to DMEM with 2% serum just before treatment for one hour (see below). After treatment, the medium was changed to DMEM 10% FBS. The cells were divided into 3 pools, namely normal, hypoxia, and hypoxiaADM, discussed in the main paper. Here, we include more details about the treatment of the cells in these pools.

Normal conditions (control): These cells were kept under normal cell-culture conditions at $37°C$ with about 21% oxygen and 5% CO_2.

Hypoxia: These cells were subjected to hypoxia (deficiency of oxygen; 0.3% O_2 level) by incubation in a hypoxic cell incubator for 60 minutes.

Hypoxia and ADM: These cells were subjected to hypoxia like the cells above, but were simultaneously treated with the peptide hormone Adrenomedullin (ADM) at 10^{-6} M concentration. This hormone is found to exhibit protective functions under various pathological conditions like ischemia in heart cells during myocardial infarction.

The cells were labelled using the live-cell friendly fluorescent marker mCLING-ATTO647N right before imaging using a concentration of 1:2000 with a 12 min incubation time. After incubation, the medium was replaced with cell-culture medium (DMEM 10% FBS) for time-lapse microscopy at $37°C$, atmospheric oxygen (i.e., the cells in hypoxia and hypoxia ADM pools are no longer in oxygen-deficient condition) and 5% CO_2. The membrane marker was quickly internalized by the cells, and labeled small membrane-bound vesicles in the cells. This membrane marker exhibits fluorescence emission maximum at 662 nm wavelength. The mitochondrial marker was imaged sequentially in a separate color channel (emission maximum at 610

nm) and used as a reference for cellular morphology. It was not a subject of computational analysis in this work.

6.2. Microscopy and imaging

GE DeltaVision Elite microscope was used for imaging the cells. The numerical aperture of the microscope is 1.42. For the membrane marker, marking the vesicles, the emission wavelength is 662 nm. Corresponding color filter was used in the collection path to image vesicles so that the light from mitochondria does not create cross-spectral noise while imaging vesicles. The digital resolution of the microscope, given by the camera pixel size divided by the magnification, is 80 nm. The exposure time for imaging the vesicles was 10 ms. The acquisition rate was 50 frames per second. Although the mitochondria are also labelled and imaged, they are not used in the analysis since they are not vesicle like structures.

7. Details of physics based simulation

Distribution of fluorescent molecules: Thousand emitters are distributed on the surface of the vesicle. Their x and y coordinates are selected using uniform distribution within the surface of the vesicle.

Photokinetics: Statistical constants used for photokinetic simulations are the same for all the molecules. A molecule may bleach (thereby become unavailable for fluorescence) with a probability 0.1. If it has not bleached, intermittent emission of fluorescence with blinking time constant of 1 milliseconds and the probability of the molecule fluorescing is 0.3. The number of emitted photons is selected from Poisson distribution with mean 100000 photons per second. These photokinetic properties and simulations have been taken from [1] and references therein.

Motion simulation: All the motion parameters that are selected randomly for any motion state assume uniform distribution in the range of the parameter. In the case of *RCir*, we first compute a random speed in the range $[0, 500]$ nm/frame assuming uniform distribution. Next, we compute a direction of update assuming uniform distribution. Using the speed and direction, new coordinates for the center of vesicle are computed and it is checked if the new coordinates are inside the circle of random motion or not. If not, the new coordinates are dropped and the process is repeated till the new coordinates are within the circle of random motion. In the case of *Rand*, such check is not performed and the new coordinates computed the first time are accepted.

Point spread function (PSF) of the microscope: The PSF of the microscope is given by an Airy function [8] as:

$$I(x_p, y_p) = \gamma\left(\frac{\mathrm{J}_1(2\pi \mathrm{r})}{2\pi r}\right)^2 \tag{3}$$

where γ is a constant related to numerical aperture (NA) of the microscope objective, magnification (M) of the microscope, refractive indices of the sample and the microscope objective media, and the emission wavelength λ. $\mathrm{J}_1(\bullet)$ is Bessel function of the first kind and the first order. Further $r = \frac{NA\hat{r}}{\lambda}$, where $\hat{r} = \sqrt{(x_p - x_s)^2 + (y_p - y_s)^2}$, (x_s, y_s) are the coordinates of a fluorescent molecule in the sample space (in true physical coordinates) and (x_p, y_p) are the effective pixel coordinates, i.e. coordinates of the pixel adjusted to match the sample size (actual physical coordinates in the camera space divided by M). The PSF is used therefore to compute an image per fluorescent molecule. The images of all the fluorescent molecules are summed up to form one microscopy image of the fluorescent molecule distribution of the sample.

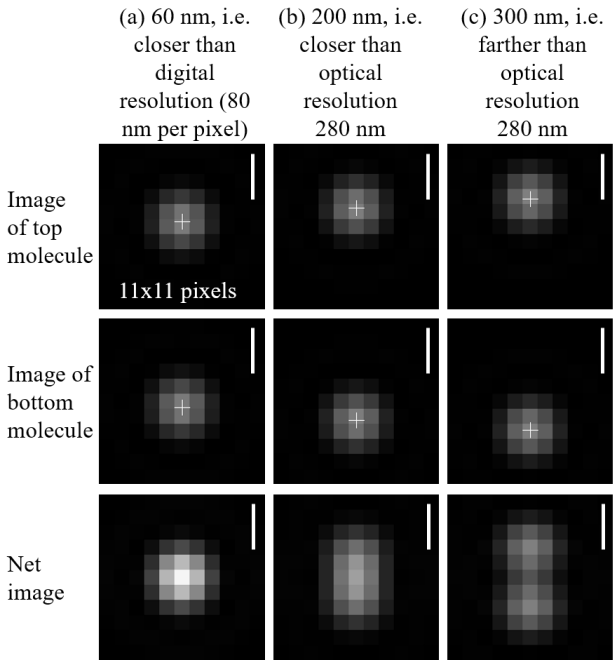

Figure 3. Illustration of the effect of digital and optical resolution of the microscope is shown through 2 fluorescent molecules, separated by different distances in (a-c). The '+' marker shows the location of emitter. The parameters used for simulation correspond to the microscope used by us. Scale bar : 500 nm.

In order to provide an insight into the optical resolution and digital resolution of the microscope, we present an example in Fig. 3. The digital resolution of the microscope is defined by the effective pixel size (80 nm in our case) and the optical resolution limit is given by Abbe as $\sim 0.61\lambda/\mathrm{NA}$ [8] (280 nm in our case).

Emulating motion through multiple time instances per frame and forming image frames: Motion occurs in continuous time coordinates while camera integrates all the temporal occurrences in the duration of one frame into a single image. We therefore simulate 10 time instances per

frame towards emulating the motion continuity. An example is presented in Fig. 4. Therefore, for simulating the raw video comprising to 200 frames for one vesicle, we form the coordinate list of all the fluorescent molecules on the vesicle for 2000 time points. Then, we compute the microscopy images for all 2000 time points. We perform temporal summation of the first 10 images to form the first frame, the next 10 images to form the second frame, and so on.

Emulating noise in the image: In microscopy images, there are two main sources of noise: (a) camera's dark current noise, also called electronic noise, which contributes a noisy background in the image, and (b) photon noise, which is based on Poisson statistics of arrival of photon at the expected location. To emulate these noise characteristics and match the signal to noise ratio observed in practical microscopy images, we perform the following operations. Let us say that the simulated microscopy image, scaled to span $[0, 1]$ is $\mathbf{I}$. Moreover, let the signal to noise ratio be SNR and the measured background values in the camera with closed shutter be b. We simulate an image $\hat{\mathbf{I}}$ as:

$$\hat{\mathbf{I}} = b(\text{SNR} - 1)\mathbf{I} + \mathbf{b} \tag{4}$$

Then, the final image with noise $\tilde{\mathbf{I}}$ is such that each pixel in $\tilde{\mathbf{I}}$ is generated using a Poisson distribution with mean equal to the corresponding pixel in $\hat{\mathbf{I}}$.

With $\sim$ms exposure time such as used in our experiments, the dark current noise is significantly stronger than the photon noise. In such situation, signal to background ratio (SBR) is a practical measure of noise. The SBR of our live-cell images are in the range $[2, 4]$. We emulate the same by randomly choosing SBR in this range ans substituting it for SNR in eq. 4. Further, we measure the mean intensity in an image acquired by shutting the aperture of the camera (i.e. in the condition when no light enters the camera). We use this value as b, which is in the range $[50, 80]$ for our microscope.

Note on 3D aspects of motion, sample, and microscope: We note that our simulations assume a 2D geometry of sample and motion as well as that the sample plane is in focus of the microscope. For a furthermore realistic physics-based modeling, we have to emulate

- 3D geometry and motion of vesicle, including their potential movement into our-of-focus region as well as region outside the depth of field of the microscope.
- 3D point spread function of the microscope such as using Gibson Lanni model [3].

8. Particle tracking challenge dataset [2] and difference from our dataset

A simulated particle dataset (PTC) is proposed in [2]. We note that PTC is quite distant from the physics of our problem. **A)** Particles in PTC displace by a few pixels between two consecutive frames. In our case, a vesicle may displace by sub-pixel distance between the two frames, see Fig. 5. **B)** PTC simulates motion as frame-to-frame displacement (1 position/frame). We simulate motion at sub-frame temporal scales (10 positions/frame), which is more physically accurate, see Fig. 6. However, it introduces motion blur in each frame, which does not help conventional tracking. **C)** Our images have small exposure time of few milliseconds. At this rate, each fluorescent molecule cannot be modeled as having a constant fluorescent intensity. Therefore, we need to include the photo-kinetics of fluorescence, which results in intensity fluctuations over time irrespective of motion. This effect is not simulated in PTC, which simplifies tracking.

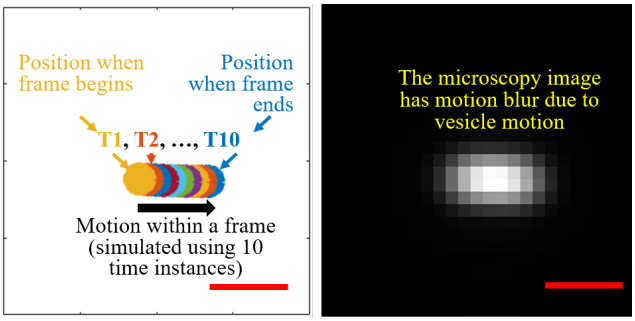

Figure 4. Multiple time instances are simulated within the duration of a single frame in order to emulate the realistic motion blur introduced by the microscope due to the motion of the vesicle. Scale bar: 500 nm.

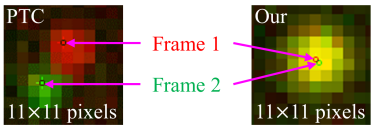

Figure 5. Our dataset often has sub-pixel displacement between two frames, while PTC has displacement of few pixels.

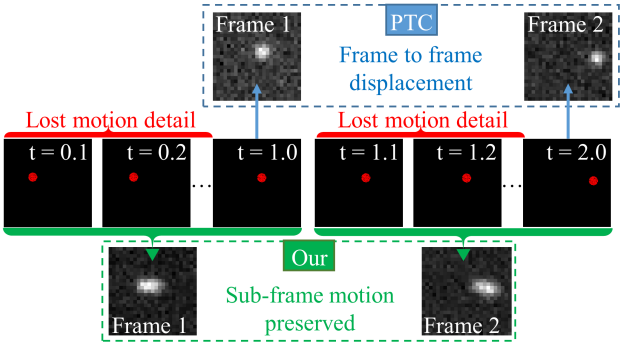

Figure 6. Our dataset preserves sub-frame motion while PTC simulates only frame to frame displacement.

9. More visual results of motion classification

A few randomly selected examples of simulated motion patterns are presented in Fig. 7 and a few random samples taken from experimental data that are correctly classified is presented in Fig. 8.

References

[1] K. Agarwal and R. Macháň. Multiple signal classification algorithm for super-resolution fluorescence microscopy. *Nature Communications*, 7:13752, 2016. 4

[2] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, et al. Objective comparison of particle tracking methods. *Nature methods*, 11(3):281, 2014. 5

[3] S. F. Gibson and F. Lanni. Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *Journal of the Optical Society of America A*, 6(9):1357–1367, 1989. 5

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1, 2

[5] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016. 1

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012. 2

[7] M. D. McDonnell and T. Vladusich. Enhanced image classification with a fast-learning shallow convolutional neural network. In *IEEE International Joint Conference on Neural Networks*, pages 1–7, 2015. 2

[8] L. Novotny and B. Hecht. *Principles of Nano-optics*. Cambridge university press, 2012. 4

[9] B. Shahriari, A. Bouchard-Côté, and N. Freitas. Unbounded bayesian optimization via regularization. In *Artificial Intelligence and Statistics*, pages 1168–1176, 2016. 2

[10] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017. 1, 2

[11] X. Xia, C. Xu, and B. Nan. Inception-v3 for flower classification. In *International Conference on Image, Vision and Computing*, pages 783–787, 2017. 2

[12] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 842–850, 2015. 2

[13] X. Zhang, J. Zou, K. He, and J. Sun. Accelerating very deep convolutional networks for classification and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):1943–1955, 2015. 2

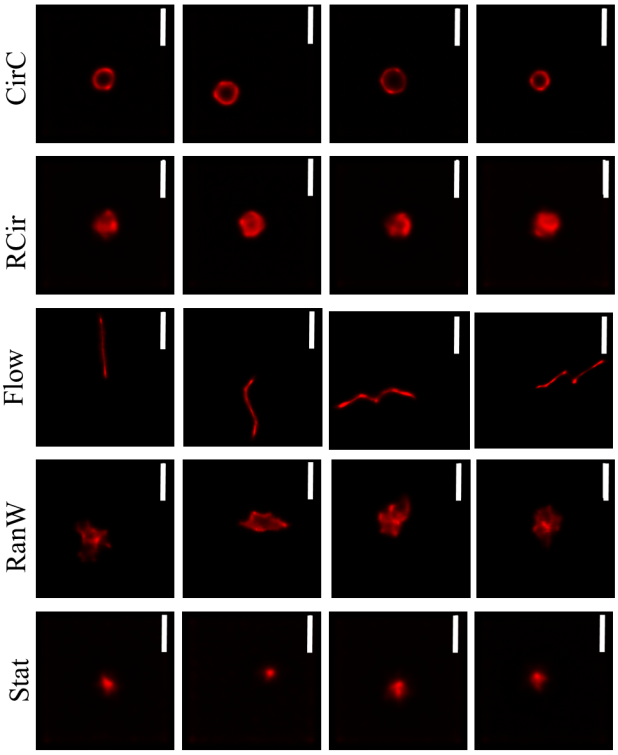

Figure 7. Few randomly selected examples of simulated motion patterns. Scale bar: 500 nm.

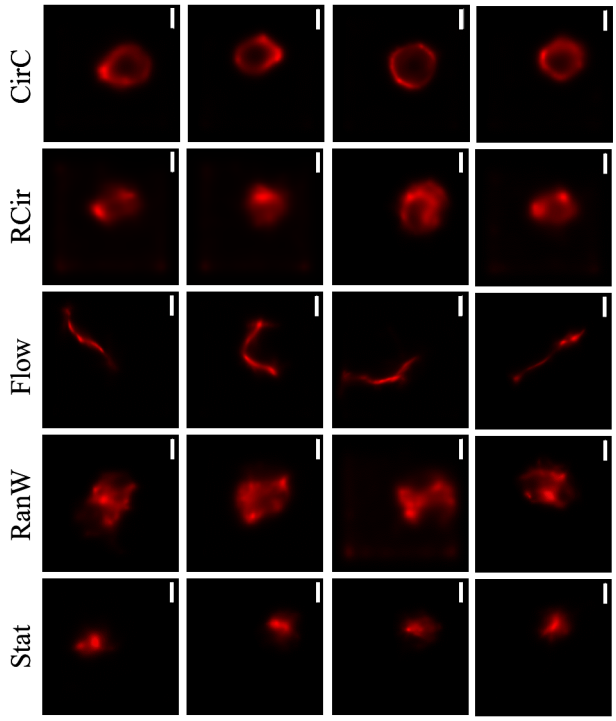

Figure 8. Few randomly selected examples of the correctly classified motion patterns. Scale bar: 500 nm.