# 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

## Abstract

*Detecting and analyzing nanoscale motion patterns of vesicles, smaller than the microscope resolution (~ 250 nm), inside living biological cells is a challenging problem. State-of-the-art CV approaches based on detection, tracking, optical flow or deep learning perform poorly on this problem. We propose an integrative approach built upon physics-based simulations, nanoscopy algorithms and shallow residual attention network to permit for the first time analysis of sub-resolution motion patterns in vesicles, also of sub-resolution diameter. Our results show state-of-the-art performance, 89% validation accuracy on simulated dataset and 82% testing accuracy on an experimental dataset of images of living heart muscle cells grown under three different pathophysiologically relevant conditions. We demonstrate automated analysis of the motion states and changes in them for over 9000 vesicles. Such analysis will enable large scale biological studies of vesicle transport and interactions in living cells in the future.*

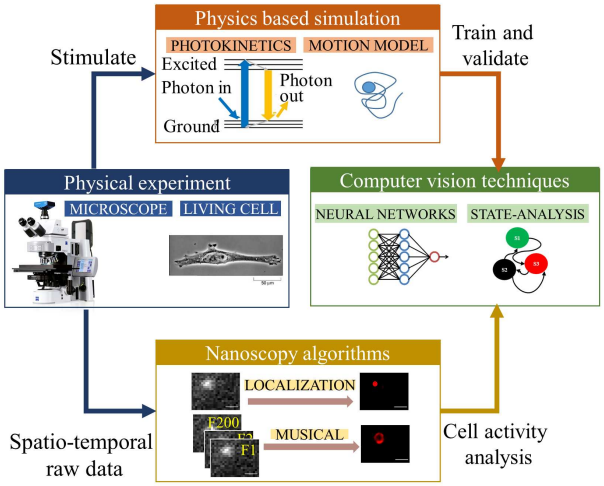


Figure 1. Our integrative approach of experiments, physics, nanoscopy, and computer vision allows analysis of nanoscale motion patterns of vesicles inside living cells.

## 1. Introduction

Microscopy images and videos are the only visual windows to the life in biological cells. The life events in a cell are orchestrated by a variety of organelles, such as nanoscale vesicles (30 nm to ~1 µm). The vesicles perform their tasks by undergoing diverse motions in the scale of tens of nanometers to a few micrometers and interacting with other sub-cellular structures. The analysis of dynamic behaviour of vesicles may hold key to understanding and treating diverse neurological and immunological disorders [21, 27, 35]. However, learning about their motion patterns from microscopy videos of vesicles inside living cells is an imposing task, both visually and through computer vision (CV), for multiple reasons presented next:

• **Optical and digital resolutions** – The digital resolution (effective pixel size) of the most advanced live-cell compatible fluorescence microscopes are limited to ~100 nm and their optical resolution (smallest resolvable feature size) is ~250 nm. As a consequence, the structures as well as the motion patterns of nanometer scale (< 250 nm) are not discernible by the microscopes, unless super-resolution microscopy (i.e. nanoscopy) approaches are employed.

• **Noise** – As compared to conventional imaging and videography, fluorescence microscopy deals with light of the order of a few photons per pixel. The shot noise and the dark noise of the camera often make the measurements significantly noisy. This has further negative effect on identification of motion patterns from microscopy videos.

• **Lack of data** – Live-cell experiments are not quite repeatable. Small variations in cell culture and imaging processes introduce differences in cell behaviour. Further, the age of the cells and the number of times of cell culture result in variations in the frequencies of normative life-events. Moreover, generating ground truth for such data is practically impossible. Therefore, generating large, controlled, statistically consistent, and suitably annotated

dataset for machine learning is quite challenging.

• **Number of vesicles and variety of motions** – A single living cell can easily contains a few hundred vesicles within the focal region of the microscope. Their diameters have a large range (30 nm to $\sim 1\,\mu m$) and motion patterns have a large variety and complexity. Designing a method that caters to such diversity is challenging.

We present an integrative approach of physics-based nanoscopy-integrated artificial intelligence for learning motion patterns of vesicles in the biological system under consideration (see Fig. 1). Our approach addresses the aforementioned problems using four key propositions.

- The complex motion patterns of individual vesicles are broken down into piece-wise simple patterns. Small spatio-temporal regions of interest (ROIs), each potentially containing a simple motion pattern of a single vesicle are identified using a combination of localization nanoscopy and particle tracking.
- Vesicles' nanoscale motion patterns smaller than the microscope resolution are reconstructed using a motion-preserving live-cell compatible nanoscopy algorithm.
- Sufficiently large annotated dataset for CV is created synthetically for diverse simple motion patterns of vesicles with a wide range of diameters using a physics-based simulation approach which emulates physical motion, fluorescence photo-kinetics, optical properties of the microscope, as well as noise. This is significantly more advanced than the previous state-of-the-art simulated vesicles' dataset [8], as discussed in the supplementary.
- A shallow residual attention network is used for learning the relatively small information content (the type of motion pattern) from a large motion-encoded nanoscopy image (hundreds of thousands of pixels for every vesicle).

We show that our approach provides significantly better results than the state-of-the-art spatio-temporal CV approaches on true microscopy videos of vesicles in heart muscle cells (cardiomyoblasts). We demonstrate that the motion patterns can be analyzed and that meaningful analytics can be derived using our approach. This analysis and the corresponding datasets is the first such contribution to the family of CV for microscopy-related research problems.

## 2. Related work

We note two separate bodies of related work. The first one pertains to the microscopy community, which is increasingly adopting CV for a variety of tasks. The second one pertains to analogous problems in CV where motion patterns of individual entities are learnt. We discuss also how our approach bridges the gaps between them.

**CV in microscopy:** Advances in microscopes and computational hardware are expanding the possibilities for live-cell image analysis, which is of importance to research in biology. Deep neural networks [50, 55] are used for tracking of cells or simulated particles. Detection based tracking [49] and feature tracking [36, 40] were successfully applied in cell migration analysis [26]. For vesicles larger than the microscope resolution, tracking and activity analysis of vesicles have been performed using single-particle tracking [8, 38, 45, 51]. Zhao et al. [58] proposed an analysis of large scale and collective motion of lysosomes (a type of vesicles) by tracking. Feature tracking works fine when particles move continuously and the signal-to-noise ratio (SNR) is high. Detection based tracking performs well when the object being tracked is a few times larger than the microscope resolution. Neither condition is satisfied in our problem. Recurrent neural networks have been used to classify spatio-temporal events [34]. Optical flow guided event detection has been applied in live-cell analysis [10]. These methods reflect promising results regarding temporal activity analysis from microscopy videos of live-cells. However, *they inherently assume that the structures and motion patterns are larger than the microscope resolution.*

**Motion pattern analysis in computer vision:** Video analysis for understanding crowd patterns [39], monitoring traffic [46], and event detection [18] are gaining popularity. They are equivalent to collective motion pattern analysis [58], single-particle tracking [38, 45], and interaction detection [51], respectively. Alexander et al. [3] introduced a computational sensor for 3D velocity measurement using a per-pixel linear constraint composed of spatial and temporal image derivatives. The challenges are however different when the sub-resolution nanoscale motion patterns in the presence of significant noise have to be investigated. Recently, micro-motion analysis [6, 13] has been proposed to extract small motion from videos that can not be observed with the naked eye. The method has been applied for extraction of micro expressions [24]. We found that these methods are sensitive to noise and therefore have limited applicability in our problem. Kim et al. [22] proposed a method for classifying human-car activity using simulated data for training. This is analogous to our approach of physics-based simulations for training. Baradel et al. [4] proposed a framework for causal learning of dynamics in mechanical systems from visual input. This is roughly analogous to our investigation of transition of vesicles from one simple motion state to another.

**Gaps bridged by our work:** The main challenge of identifying nanoscale motion patterns is solved by selecting a motion-preserving nanoscopy algorithm, namely multiple signal classification algorithm (MUSICAL) [1], for performing optical and digital super-resolution for live-cell imaging. Through this, we introduce live-cell compatible nanoscopy algorithms [1, 9, 12, 42] as valuable tools for CV at the nanometer scale. Although analysing nanoscopy images using neural networks may help in various biological experiments, the application of state-of-the-art deep learn-

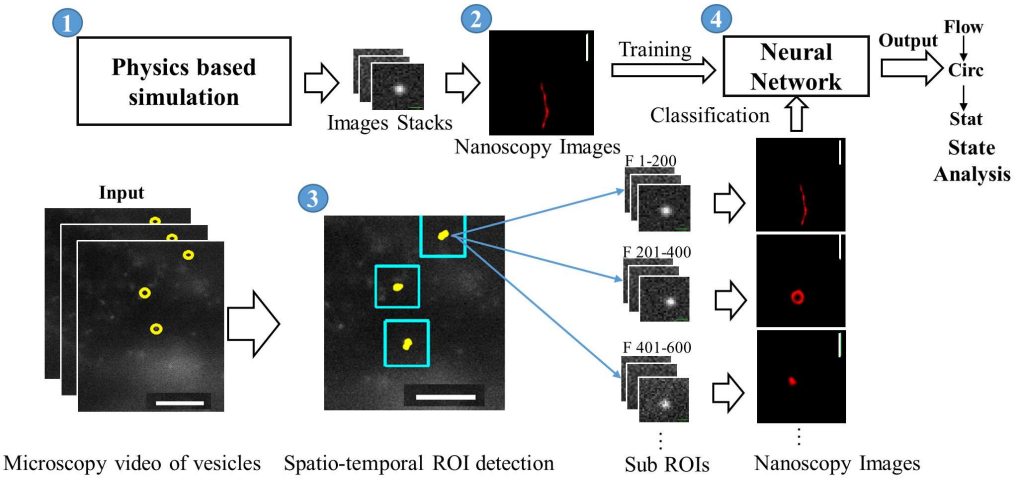


Figure 2. Overview of the proposed framework. Scale bars: 5 μm horizontal, 500 nm vertical. F indicates frame number hereon.

ing methods for nanoscopy image analysis is limited. This, we opine, is due to 1) limited availability of annotated large datasets, and 2) presence of fewer features in nanoscopy images compared to the real-world camera images. The first problem is solved in our case by employing a rigorous physics-based simulation framework which emulates both the dynamic organelles and the presence of noise in the experiments. All details of the physics-based simulations are included in the supplementary. For problems in biology where ground truth on experimental data is nearly impossible, such approaches will be indispensable for developing CV solutions. Such approach will also find value in other ground-truth deficient applications such as astronomy, geology and climate if suitable physics-based simulation frameworks of sufficient detail can be developed. The second problem is solved by using a shallow residual attention network. The features exploited in state-of-the-art deep models based CV, namely textures, edges, and colors, are missing in the microscopy data. Moreover, the dynamic range of intensity is quite small in microscopy images and the noise is comparable to the signal. The microscopy images contain only few features encoded mainly in intensity variations. Due to these reasons, we expect shallow networks to perform better than deep models. This introduces a valuable CV tool to the microscopy community, which currently depends heavily on visual inspection.

## 3. Method

The proposed methodology is shown in Fig. 2. It consists of four modules: (1) physics-based simulations for creating training dataset, (2) MUSICAL for nanoscale motion reconstruction, (3) spatio-temporal ROI detection using localization based tracking, and (4) classification of motion patterns. We discuss each module next.

### 3.1. Physics-based simulations

Our simulation flowchart is shown in Fig. 3(a). We first simulate a vesicle labeled with several fluorescent molecules. The diameters of the simulated vesicles is in the range $[150, 400]$ nm. The fluorescent molecules are randomly placed inside the volume of the vesicle. The number of photons emitted by each molecule are simulated using the photokinetic model of [1]. Code provided by its authors used for this. It includes blinking, bleaching, and non-radiative energy dissipation of fluorescent molecules [9]. It has been reported that the vesicles may demonstrate random movement in a confined space [2], directed flow-like motion [7], circular motion [32], and sometimes they become stationary during interaction with other organelles [14]. Inspired by the biological evidence, we have simulated five types of vesicular motion patterns (also called motion states) in 2D, described below:

- **Circular Motion (*Circ*):** The vesicle moves along the periphery of a virtual circle with randomly selected center, radius, and velocity. The radius of the circle and the velocity of the vesicle are in the ranges $[200, 500]$ nm and $[0, 500]$ nm/frame, respectively.
- **Random walk inside a circle (*RCir*):** The vesicle takes random positions within a circular area. The radius of the circle is chosen randomly from the range $[200, 400]$ nm.
- **Flow (*Flow*):** The vesicle moves along a path with a constant velocity. First, a random curve is generated. Next, the vesicle is transported along the curve with velocity selected randomly from the range $[0, 1000]$ nm/frame.
- **Random walk (*RanW*):** During a random walk, the vesicle may move in any direction with equal probability. For each movement, the velocity is randomly selected from the range $(0, 1000]$ nm/frame.
- **Stationary (*Stat*):** The vesicle remains stationary.

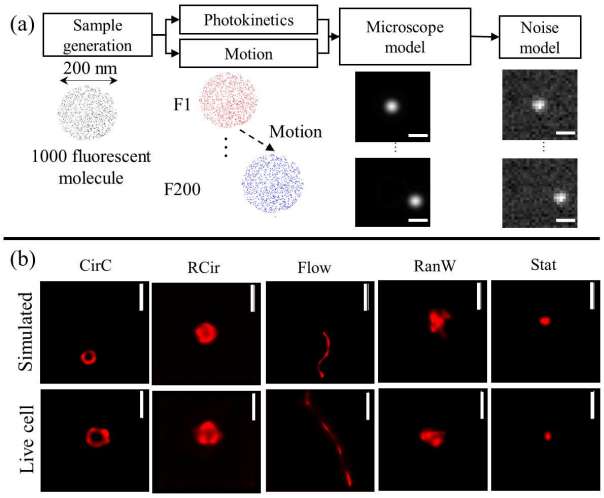


Figure 3. Physics-based simulation framework. (a) The flow chart and its illustration using an example of a vesicle of diameter 200 nm. (b) A visual comparison of a few randomly selected examples of the chosen motion patterns. Scale bar: 500 nm.

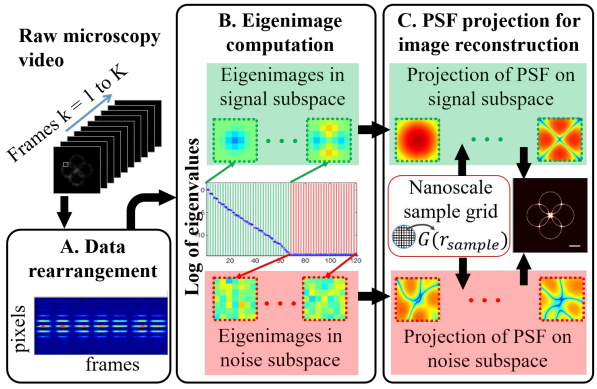


Figure 4. MUSICAL preserves spatio-temporal features in images using eigenimages (block B) and reconstructs the nanoscale patterns by projecting microscope's PSF from a nanoscale sample grid onto the signal and noise subspaces (block C).

We note that our library of motion patterns is not exhaustive. It is expandable to include other patterns in the future. After forming the coordinate list of all the fluorescent molecules at all the time points, we compute the raw noise-free microscopy video by emulating the point spread function (PSF) [31] using the optical parameters relevant to the molecules, the microscope, and the imaging conditions. Then, the noise characteristics of the camera are incorporated [44]. All the details are included in the supplementary. We show an example of simulation below the block diagram presented in Fig. 3(a). We also illustrate examples of simulated motion patterns reconstructed using MUSICAL as compared to similar reconstructions from the experimental live-cell data in Fig. 3(b).

### 3.2. MUSICAL

The function of MUSICAL [1] is explained in two parts, namely eigenimages and identifying nanoscale patterns.

**Spatio-temporal features in eigenimages:** For small optical windows (size given by the span of the microscope PSF), MUSICAL computes eigenimages from the microscopy video. The eigenimages order the spatio-temporal information from the most consistent ones to most random ones. The first few eigenimages with largest eigenvalues correspond to vesicle motion patterns (spanning the signal subspace) and the remaining correspond to noise patterns (spanning the noise subspace), see Fig. 4.

**Nanoscale pattern identification:** Even if two points are separated by a distance below both the optical and the digital resolution, the PSFs at such points are slightly different from each other. Their projection onto the signal and noise subspaces are therefore different. Precisely, at a point in the sample space, the projection of the PSF onto every single eigenimage in the noise subspace is zero if two conditions are satisfied. First, the separation of signal and noise subspaces is robust. Second, a fluorescent molecule ever emitted fluorescence photons from that location during the video. The condition of zero projection on the noise subspace is violated at a point even slightly away from such a location. This property is mathematically enhanced in MUSICAL to reconstruct nanoscopy image with pronounced nanoscale features.

### 3.3. Spatio-temporal ROI detection

This step comprises of two tasks - detecting vesicles and linking the detections across frames (Fig. 5).

**Detection of vesicles:** Localization nanoscopy [41] can localize individual fluorescent molecules by fitting Gaussian functions in microscopy images. This is possible only if extreme spatio-temporal sparsity in fluorescence emissions is enforced, which is not possible while imaging living cells. Nonetheless, the nearly spherical geometry of vesicles implies that their image can also be roughly approximated as a Gaussian functions. Thus, we use localization nanoscopy in an unconventional setting for detecting vesicles in the microscopy videos. We have used quickPALM [17] implementation for this purpose.

**Linking the detections and creating sub ROIs:** The detected vesicles are linked using Hungarian method and Kalman filter [5] to construct their trajectories. Let a given live-cell sequence contain $n$ number of tracks as: $\{T_1, T_2, ..., T_n\}$. Each track is defined by series of positions of the vesicle over time, i.e. $\{p_1, p_2, ..., p_m\}$, where $p_i = (x_i, y_i)$. For each track, a set of sequential non-overlapping sub ROIs is created such that each sub ROI contains $\hat{K}$ continuous positions of the particle. The key idea behind using sub ROIs is that each sub ROI is likely to contain one simple motion pattern, potentially among *Circ*, *RCir*, *Flow*,

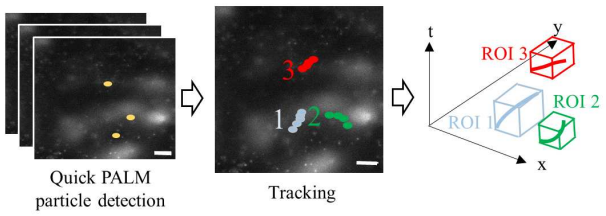


Figure 5. ROI detection using localization based tracking.

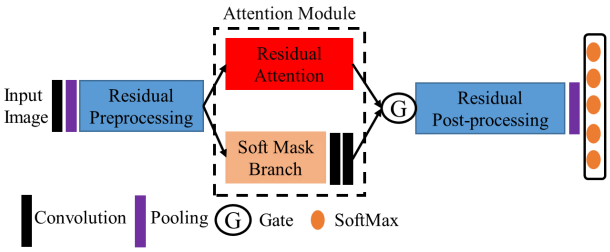


Figure 6. Architecture of the shallow residual attention network.

*RanW*, and *Stat*. The number $\hat{K}$ can either be selected for the chosen biological cell type and image acquisition rate heuristically or more sophisticated automatic sub ROI selection may be designed, which is out of the scope of the current work. We have heuristically selected $\hat{K} = 200$.

### 3.4. Motion Classification

The choice of the depth of the network depends on the task, image features, and class variation. Several biological classification tasks have been solved using shallow networks [11, 15, 33] due to the unavailability of large microscopy and nanoscopy datasets as well as fewer features in live-cell images compared to the real-world RGB images. We have observed that the state-of-the art deep neural networks such as deep CNN [23], VGG16 [56], Inception [52], and ResNet50 [16] performed poorly in our dataset (results in section 4). Furthermore, the use of pretrained models did not improve the classification accuracy significantly. We found that comparatively shallow networks such as a 3-layered MLP, shallow CNN [28], and ResNet20 perform better on our data. The observations inspired us to design a shallow network for motion pattern classification.

In the last few years, the use of residual connection among layers has proven its ability to improve accuracy in several computer vision tasks [16]. On the other hand, attention-based neural networks inspired by the human perception have become popular in various computer vision tasks. They employ attention mechanism [53] to identify and highlight useful features during learning. Recently, residual-attention mechanism [47] demonstrated state-of-the-art or comparable accuracy in certain computer vision tasks [20, 30, 57], and also serve as an inspiration for us.

**Shallow Residual Attention Network:** We combine the concept of residual and attention mechanisms with a shallow neural network to propose a Shallow Residual Attention Network (SRAN). The network architecture is presented in Fig. 6. It consists of a set of initial pre-processing layers including a residual pre-processing block, an attention module, and a gated residual post-processing block connected to the classification layer. The attention module further consists of a residual attention block (also called trunk branch) and a soft mask branch. The trunk branch has a down-sample and an up-sample unit, for top-down and bottom-up attention mechanisms [47] respectively. The soft mask branch is a form of residual block. The outputs of the trunk and soft mask branches are combined using a controlled gate similar to long short-term memory. The attention module suppresses the noise and highlights important information by applying dot product between the residual attention features and soft masks learnt in the trunk branch and the soft mask branch respectively. The details of SRAN are given in the supplementary.

## 4. Experimental results

### 4.1. Dataset

In order to evaluate the effectiveness of the proposed method, we use two datasets described below. We make both the datasets and supplementary public for research purposes at our project page[1].

**Simulation dataset:** This dataset is used for training and evaluation of the classifier. It contains 3000 data samples for each type of motion pattern. Each data sample is a small video of 200 frames corresponding to simulated microscopy images of 25 × 25 pixels of a single vesicle exhibiting a single motion pattern. The optical and camera parameters used for the simulation were based on the experimental setup used for creating live-cell dataset. The simulated noise was chosen such that the signal to noise ratio was similar to the videos in the live-cell dataset.

**Live-cell dataset:** Cardiomyoblasts (heart muscle cells) were divided into 3 different pools and labelled using live-cell friendly fluorescent dye. The pools are: • **Normal:** These cells were kept under normal cell-culture conditions. • **Hypoxia:** These cells were subjected to hypoxia (deficiency of oxygen) for 1 hour. • **HypoxiaADM:** These cells were subjected to hypoxia like the cells above, but were simultaneously treated with the hormone adrenomedullin (ADM). This hormone is found to exhibit protective functions under pathological conditions like myocardial infarction (cardiac arrest).

For each pool, 10 videos of 2000 frames each and 1024 × 1024 pixels were imaged using GE DeltaVision Elite fluorescent microscope. Other experimental details are provided in the supplementary. We counted the number of vesicles in the cells that were imaged in each pool. These

[1] https://nonoscalemotion.github.io/

Table 1. Multiple Object Tracking Accuracy [19] of different methods on live-cell dataset.

| Condition | Feature Tracking [40] | Deep Tracking [49] | Proposed |
|---|---|---|---|
| Normal | 0.48 | 0.69 | 0.91 |
| Hypoxia | 0.39 | 0.62 | 0.93 |
| HypoxiaADM | 0.41 | 0.68 | 0.87 |

Table 2. Classification accuracy of different neural networks using various input features. Format: Validation/Testing

| Method | Raw Images | Micro Motion | Optical Flow |
|---|---|---|---|
| RNN [29] | 0.29 / 0.26 | 0.26 / 0.24 | 0.32 / 0.21 |
| BLSTM [25] | 0.32 / 0.21 | 0.27 / 0.18 | 0.36 / 0.24 |
| Con3D [54] | 0.28 / 0.26 | 0.22 / 0.22 | 0.46 / 0.39 |

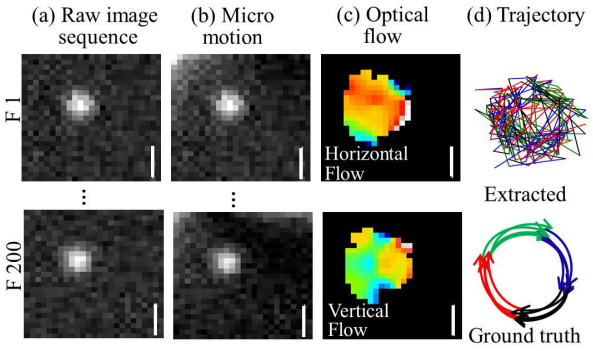


Figure 7. Feature representations of a vesicle in *Circ* state using different approaches for motion classification. In (d), each colour represents different direction quadrant. Scale bar: 500 nm.

numbers are 3283 vesicles for normal, 3186 vesicles for hypoxia, and 2980 vesicles for hypoxiaADM. Thus, we performed activity analysis of experimental data of a total of 9449 vesicles. The motion patterns of sub ROIs of each vesicle were manually annotated for generating ground truth by visual inspection of raw image sequences and nanoscopy images reconstructed using MUSICAL. Live-cell dataset refers to all the data, except in section 4.5 4.4 where pool-specific results are presented.

### 4.2. Vesicle Localization and Tracking

We experimented with feature tracking [40], deep learning based tracking [43], and the proposed localization based tracking. In deep learning based tracking, the neural network was trained with the simulated dataset and tested on live-cell dataset. We evaluated the tracking performance using multiple object tracking accuracy (MOTA) [19] metric with manually generated ground truth, see results in Table 1. Feature based tracking method failed to distinguish between features and noise, therefore failing to track. Deep learning based tracking methods also perform poor due to noise and tiny size of the vesicles.

### 4.3. Results of Motion Classification

We conducted different experiments using a variety of spatio-temporal features and learning methods. We tried using raw image sequences, micro-motion magnified sequences [13], optical flow, and the trajectories constructed in the proposed ROI detection approach as the input for classification. Fig. 7 depicts a visual comparison of the different features extracted for a vesicle in *Circ* state. It can be observed from Fig. 7 that the naked eye can not detect the *Circ* pattern from either the raw image sequence or the micro-motion magnified sequence (example in the supplementary videos). The micro-motion magnified sequence contains larger noise compared to the raw image sequence. Due to high noise levels in the raw data, optical flow spans a larger area, therefore failing to detect the nanoscale motion. Localization nanoscopy can detect the vesicle but can not extract the trajectory of nanoscale movement accurately. We experimented using LSTM (baseline) and a deep CNN [48] using the detected trajectories as input and found the accuracy of (validation/testing) as (0.38/0.29) and (0.40/0.35) for LSTM and deep CNN, respectively. For the other features, namely raw image sequence, micro-motion magnified sequence, and optical flow, we experimented using different baseline learning algorithms. For all the experiments, the simulation dataset is used for training and validation. Five-fold cross-validation is used. The live-cell dataset is used for testing. Parameters of all the baseline methods are set similar to the original implementations. We have included early stopping and data augmentation , and verified that no over-fitting exists (see the supplementary for training details, hyperparameters, and hyperparameter study). The classification accuracy is presented in Table 2. The results indicate that that these features are not suitable for the classification of nanoscale movement.

Next, we performed experiments to classify the motion patterns using the nanoscopy images obtained using MUSICAL as inputs. SRAN is trained and tested with a similar weight initialization method and residual blocks reported in [47]. We used 2-stage attention block (compared to a 3 stage attention block reported in [47]); training details are in the supplementary. It took 35 epochs to stabilize the learning (see Fig. 8). In the case of the baseline methods, we keep most of the settings same as the original implementations. The results are summarized in Table 3. It is observed that most shallow networks perform better compared to deep networks and SRAN performs the best. Fig. 8 presents the comparative epoch vs accuracy and loss of a deep residual attention network [47] (DRAN) and SRAN. It is seen that SRAN stablizes and converges quicker and to a lower loss than the deep counterpart.

**Failure cases:** Fig. 9 depicts the confusion matrix of SRAN for the live-cell dataset. Although the accuracy for each individual class is better than 70%, we make some in-

Table 3. Classification accuracy of different methods using nanoscopy images. Format: Validation/Testing

| Method | Pre-training | Accuracy |
|---|---|---|
| Deep CNN [23] | Imagenet | 0.32 / 0.29 |
| Deep CNN [23] | - | 0.36 / 0.31 |
| VGG16 [56] | Imagenet | 0.42 / 0.33 |
| VGG16 [56] | - | 0.33 / 0.33 |
| Attention Model [53] | - | 0.71 / 0.56 |
| Shallow Network [28] | - | 0.82/ 0.63 |
| ResNet50 [16] | - | 0.71/ 0.69 |
| ResNet20 [16] | - | 0.82/ 0.74 |
| MLP (Baysian Optimization) [37] | - | 0.72/ 0.68 |
| Inception V3 [52] | Imagenet | 0.46 / 0.36 |
| Inception V3 [52] | - | 0.43 / 0.29 |
| Deep residual attention [47] | - | 0.85/ 0.78 |
| Proposed SRAN | - | **0.89/ 0.82** |

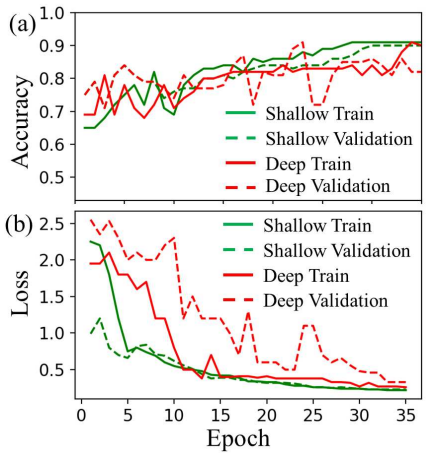


Figure 8. Accuracy & loss curves of DRAN and SRAN.

teresting observations. The miss-classifications are generally among the classes where randomness at nanoscale is involved and therefore random patterns of two kinds may have significant overlap. In other cases, artefacts due to noise in the nanoscale reconstruction may be easily confused with an equivalent nanoscale random motion pattern. In yet other cases, more than one vesicles present may be present in close vicinity, resulting in multiple motion reconstructions in a single ROI. Fig. 10 presents some failure cases related to the points mentioned above.

## 4.4. Analysis of Events

We analyzed the frequency of motion patterns and changes in motion patterns (i.e. events) in the live-cell dataset. Fig. 11(a) shows the statistics of motion states in normal, hypoxia, and hypoxiaADM pools. A clear demarcation is observed between them, except for the *Stat* motion state. Here, we see that vesicles in the case of hypoxia are least stationary. Potentially, adding ADM restores the occurrence of vesicles in this state towards normal pool.

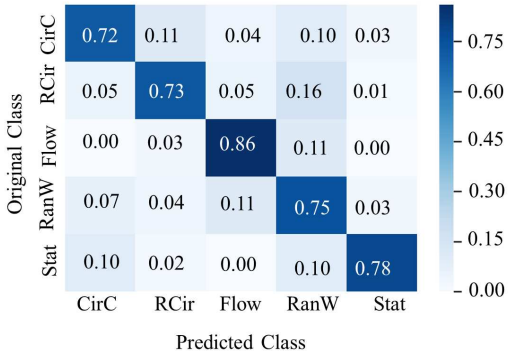


Figure 9. Confusion matrix on the live-cell dataset using SRAN.

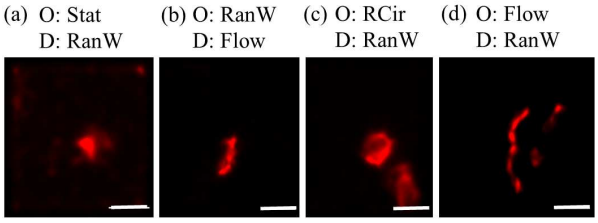


Figure 10. Example failure casess. D: detected, O: ground truth.

We also note that most vesicles in any pool are in the *RanW* state. Fig. 11(b) shows the statistics of changes in motion states in normal, hypoxia, and hypoxiaADM pools. It is of particular interest to note the squares with green background. They indicate that ADM may have resulted into change in the trend introduced by hypoxia. For example, as compared to normal pool, hypoxia pool demonstrated more number of transitions from *Circ* and *Flow* to *RanW* states. But, hypoxiaADM demonstrated reduced number of such transitions. Other similar behaviours may indicate some potential mechanisms of action of ADM. It is important to note that these results are not conclusive from biological perspective since these experiments were designed to provide an initial test dataset for the proposed framework. A rigorous biological study needs further biological and environmental controls, hypothesis-specific experiment design, and large scale experimentation.

We further show that our analysis may indicate nanoscale nature of interaction of two sub-cellular structures. For example, in Fig. 12, green colored low resolution structures are mitochondria. A vesicle flows towards it and interacts with it. This is visible in the microscopy video, included in the supplementary. However, the nanoscale detail of interaction is not known. The result of our framework, with 200 frames for each sub ROI, is presented in Fig. 12(a). The interaction is contained in sub ROI 2, which is classified as *RCir*. Then, we used the proposed framework with only 50 frames per sub ROI. This result, presented in Fig. 12(b), indicates that sub ROIs 5-8 contain the inter-

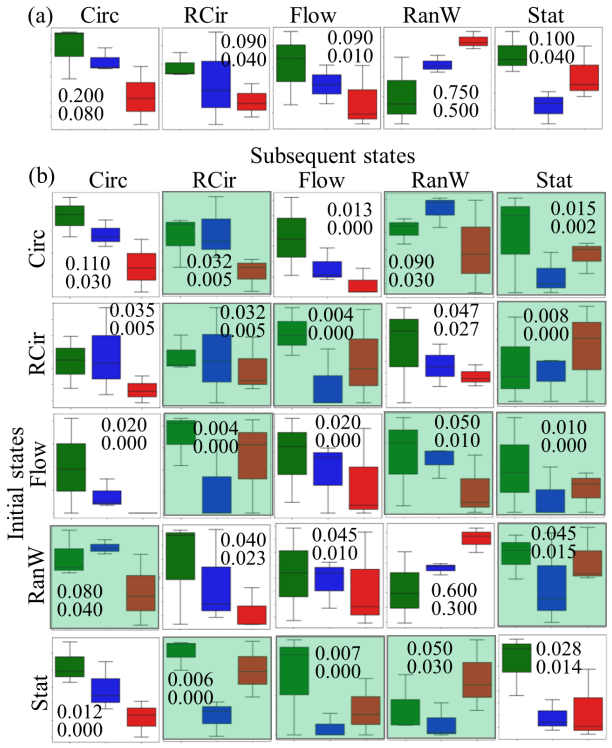


Figure 11. Analytics of motion patterns and changes in them. Legend for box plots: normal (green), hypoxia (blue), and hypoxiaADM (red). Numbers in each square indicate the maximum and minimum values for that square. (a) frequency of occurrence of motion patterns (ratio of sub ROIs in a particular motion state to the total number of sub ROIs in a pool). (b) ratio of number of consecutive-motion-state-pairs exhibiting a certain combination of initial and subsequent motion states to the total number of consecutive-motion-state-pairs. In (b), squares with green background indicate a trend reversal in hypoxiaADM as compared to trend of change between normal and hypoxia pools.

action. Among them, sub ROIs 5-7 are classified as *Stat* and generate nanoscopy spots at three different locations (see magenta, cyan, and blue spots below the white pattern) while the sub ROI 8 is classified as *Circ*. This indicates that the vesicle may have spent some time being stationary at different locations (hopping action) in close vicinity of mitochondrion, before performing a circular motion (spinning action) close to it. Such analysis will open possibilities of understanding detailed mechanisms of interactions.

## 5. Discussion and conclusion

We report a first framework and an important step towards studying motion and interaction of vesicles in living biological cells and cell systems with sub-resolution nanoscale details. Our approach indicates the utility of hybrid learning approaches which combine non-CV approaches with conventional CV approaches to perform challenging tasks with specific limitations due to the nature and physics of microscopy data. Our work also highlights that shallow learning networks may outperform deep learning networks for certain tasks where feature sparsity is an important characteristic of the data. We envision at least three future directions for the developed framework of analysis. First, the simulation framework can be extended to 3D to incorporate out of focus light and limited depth of focus of microscopes. Second, more variety of motion patterns can be incorporated in this framework or custom motion states may be learnt for different sub-cellular and inter-cellular structures. Third, the complete sequence of motion states can be formed to identify specific events of interest. The correlation of such events with activities of other sub-cellular structures can be used to identify and better understand biological interactions.

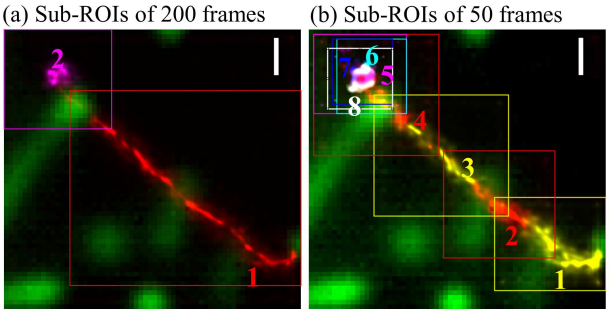


Figure 12. Example of interaction of a vesicle (nanoscopy images obtained using MUSICAL in colors other than green) with another sub-cellular structure namely mitochondrion (green microscopy image) and effect of choosing sub ROIs of different temporal sizes. In (a), sub ROIs 1 and 2 are classified as *Flow* and *RCir*, respectively. In (b), sub ROIs 1-4 are classified as *Flow*, sub ROIs 5-7 as *Stat*, and sub ROI 8 as *Circ*. Scale bars: 500 nm.

Our framework can accommodate different time scales (as demonstrated in Fig. 12) for extracting motion details of different levels. In this sense, the framework is easily adaptable to different imaging conditions. In the future, the applicability of this framework for sub-resolution analysis of microscopy images and videos from a wide variety of microscopes and biological problems will be explored.

## Acknowledgement

The following funding is acknowledged: ERC starting grant 804233 (Agarwal), Research Council of Norway's Nano2021 grant 288565 (Ahluwalia), Northern Norway Regional Health Authority grant HNF1449-19 (Myrmel and Birgisdotiir), UiT's strategic funding program (Sekh), and UiT's Tematiske Satsinger grants (all authors). All data and codes are available at https://nonoscalemotion.github.io/.

## References


[1] K. Agarwal and R. Macháň. Multiple signal classification algorithm for super-resolution fluorescence microscopy. *Nature Communications*, 7:13752, 2016. 2, 3, 4

[2] H. Al-Obaidi, B. Nasseri, and A. T. Florence. Dynamics of microparticles inside lipid vesicles: movement in confined spaces. *Journal of Drug Targeting*, 18(10):821–830, 2010. 3

[3] E. Alexander, Q. Guo, S. Koppal, S. Gortler, and T. Zickler. Focal flow: Measuring distance and velocity with defocus and differential motion. In *European Conference on Computer Vision*, pages 667–682, 2016. 2

[4] F. Baradel, N. Neverova, J. Mille, G. Mori, and C. Wolf. Cophy: Counterfactual learning of physical dynamics. *arXiv preprint arXiv:1909.12000*, 2019. 2

[5] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple online and realtime tracking. In *IEEE International Conference on Image Processing*, pages 3464–3468, 2016. 4

[6] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 105–110, 2013. 2

[7] B. Cabukusta and J. Neefjes. Mechanisms of lysosomal positioning and movement. *Traffic*, 19(10):761–769, 2018. 3

[8] 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. 2

[9] S. Cox, E. Rosten, J. Monypenny, T. Jovanovic-Talisman, D. T. Burnette, J. Lippincott-Schwartz, G. E. Jones, and R. Heintzmann. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature Methods*, 9(2):195, 2012. 2, 3

[10] A. Czirok, D. G. Isai, E. Kosa, S. Rajasingh, W. Kinsey, Z. Neufeld, and J. Rajasingh. Optical-flow based noninvasive analysis of cardiomyocyte contractility. *Scientific Reports*, 7(1):10404, 2017. 2

[11] M. R. de Souza, R. Ruschel, A. Susin, J. M. Boeira, L. V. Guimares, and A. Parraga. A framework for automatic recognition of cell damage on microscopic images using artificial neural networks. In *International Conference of Engineering in Medicine and Biology Society*, pages 636–639, 2018. 5

[12] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein. Fast, background-free, 3D super-resolution optical fluctuation imaging (sofi). *Proceedings of the National Academy of Sciences*, 106(52):22287–22292, 2009. 2

[13] M. Elgharib, M. Hefeeda, F. Durand, and W. T. Freeman. Video magnification in presence of large motions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4119–4127, 2015. 2, 6

[14] Y. Han, M. Li, F. Qiu, M. Zhang, and Y.-H. Zhang. Cell-permeable organic fluorescent probes for live-cell long-term super-resolution imaging reveal lysosome-mitochondrion interactions. *Nature Communications*, 8(1):1307, 2017. 3

[15] E. A. Hay and R. Parthasarathy. Performance of convolutional neural networks for identification of bacteria in 3D microscopy datasets. *PLoS Computational Biology*, 14(12):e1006628, 2018. 5

[16] 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. 5, 7

[17] R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga. Quickpalm: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nature Methods*, 7(5):339, 2010. 4

[18] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7842–7851, 2019. 2

[19] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):319–336, 2008. 6

[20] J.-H. Kim, S.-W. Lee, D. Kwak, M.-O. Heo, J. Kim, J.-W. Ha, and B.-T. Zhang. Multimodal residual learning for visual QA. In *Advances in Neural Information Processing Systems*, pages 361–369, 2016. 5

[21] S. Kim, Y. Sato, P. S. Mohan, C. Peterhoff, A. Pensalfini, A. Rigoglioso, Y. Jiang, and R. A. Nixon. Evidence that the rab5 effector appl1 mediates app-$\beta$ctf-induced dysfunction of endosomes in down syndrome and alzheimer's disease. *Molecular Psychiatry*, 21(5):707, 2016. 1

[22] T. S. Kim, M. Peven, W. Qiu, A. Yuille, and G. D. Hager. Synthesizing attributes with unreal engine for fine-grained activity analysis. In *IEEE Winter Applications of Computer Vision Workshops*, pages 35–37, 2019. 2

[23] 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. 5, 7

[24] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*, 9(4):563–577, 2017. 2

[25] Y. Mao and Z. Yin. Two-stream bidirectional long short-term memory for mitosis event detection and stage localization in phase-contrast microscopy images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 56–64, 2017. 6

[26] P. Masuzzo, M. Van Troys, C. Ampe, and L. Martens. Taking aim at moving targets in computational cell migration. *Trends in Cell Biology*, 26(2):88–110, 2016. 2

[27] J. M. Mc Donald and D. Krainc. Lysosomal proteins as a therapeutic target in neurodegeneration. *Annual Review of Medicine*, 68:445–458, 2017. 1

[28] 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. 5, 7

[29] A. Montes, A. Salvador, S. Pascual, and X. Giro-i Nieto. Temporal activity detection in untrimmed videos with recurrent neural networks. *arXiv preprint arXiv:1608.08128*, 2016. 6

[30] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *IEEE International Conference on Computer Vision*, pages 1520–1528, 2015. 5

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

[32] N. Okabe, B. Xu, and R. D. Burdine. Fluid dynamics in zebrafish kupffer's vesicle. *Developmental Dynamics: an official publication of the American Association of Anatomists*, 237(12):3602–3612, 2008. 3

[33] T. Pärnamaa and L. Parts. Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning. *G3: Genes, Genomes, Genetics*, 7(5):1385–1392, 2017. 5

[34] H. T. H. Phan, A. Kumar, D. Feng, M. Fulham, and J. Kim. Unsupervised two-path neural network for cell event detection and classification using spatiotemporal patterns. *IEEE Transactions on Medical Imaging*, 38(6):1477–1487, 2018. 2

[35] N. Plotegher and M. R. Duchen. Mitochondrial dysfunction and neurodegeneration in lysosomal storage disorders. *Trends in Molecular Medicine*, 23(2):116–134, 2017. 1

[36] I. F. Sbalzarini and P. Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*, 151(2):182–195, 2005. 2

[37] B. Shahriari, A. Bouchard-Côté, and N. Freitas. Unbounded bayesian optimization via regularization. In *Artificial Intelligence and Statistics*, pages 1168–1176, 2016. 7

[38] H. Shen, L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, B. Shuang, and C. F. Landes. Single particle tracking: from theory to biophysical applications. *Chemical Reviews*, 117(11):7331–7376, 2017. 2

[39] Z. Shen, Y. Xu, B. Ni, M. Wang, J. Hu, and X. Yang. Crowd counting via adversarial cross-scale consistency pursuit. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5245–5254, 2018. 2

[40] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc. Feature tracking and matching in video using programmable graphics hardware. *Machine Vision and Applications*, 22(1):207–217, 2011. 2, 6

[41] A. R. Small and R. Parthasarathy. Superresolution localization methods. *Annual Review of Physical Chemistry*, 65:107–125, 2014. 4

[42] O. Solomon, Y. C. Eldar, M. Mutzafi, and M. Segev. Sparcom: sparsity based super-resolution correlation microscopy. *SIAM Journal on Imaging Sciences*, 12(1):392–419, 2019. 2

[43] R. Spilger, T. Wollmann, Y. Qiang, A. Imle, J. Y. Lee, B. Müller, O. T. Fackler, R. Bartenschlager, and K. Rohr. Deep particle tracker: Automatic tracking of particles in fluorescence microscopy images using deep learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 128–136. 2018. 6

[44] E. T. T. T. Stephanie Fullerton, Keith Bennett. ORCA-flash4.0 - changing the game. Technical report, Hamamatsu, 2010. 4

[45] J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, and K. W. Eliceiri. Trackmate: An open and extensible platform for single-particle tracking. *Methods*, 115:80–90, 2017. 2

[46] M.-T. Tran, T. Dinh-Duy, T.-D. Truong, V. Ton-That, T.-N. Do, Q.-A. Luong, T.-A. Nguyen, V.-T. Nguyen, and M. N. Do. Traffic flow analysis with multiple adaptive vehicle detectors and velocity estimation with landmark-based scanlines. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 100–107, 2018. 2

[47] 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. 5, 6, 7

[48] L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4305–4314, 2015. 6

[49] Y. Wang, M. Ali, Y. Wang, S. Kucenas, and G. Yu. Detection and tracking of migrating oligodendrocyte progenitor cells from in vivo fluorescence time-lapse imaging data. In *IEEE International Symposium on Biomedical Imaging*, pages 961–964, 2018. 2, 6

[50] Y. Wang, H. Mao, and Z. Yi. Stem cell motion-tracking by using deep neural networks with multi-output. *Neural Computing and Applications*, pages 1–13, 2017. 2

[51] Y. C. Wong, D. Ysselstein, and D. Krainc. Mitochondria–lysosome contacts regulate mitochondrial fission via rab7 gtp hydrolysis. *Nature*, 554(7692):382, 2018. 2

[52] X. Xia, C. Xu, and B. Nan. Inception-v3 for flower classification. In *International Conference on Image, Vision and Computing*, pages 783–787, 2017. 5, 7

[53] 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. 5, 7

[54] H. Xu, A. Das, and K. Saenko. R-c3d: Region convolutional 3D network for temporal activity detection. In *Proceedings of the IEEE international conference on computer vision*, pages 5783–5792, 2017. 6

[55] Y. Yao, I. Smal, and E. Meijering. Deep neural networks for data association in particle tracking. In *IEEE International Symposium on Biomedical Imaging*, pages 458–461, 2018. 2

[56] 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. 5, 7

[57] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 286–301, 2018. 5

[58] H. Zhao, Q. Zhou, M. Xia, J. Feng, Y. Chen, S. Zhang, and X. Zhang. Characterize collective lysosome heterogeneous dynamics in live cell with a space-and time-resolved method. *Analytical Chemistry*, 90(15):9138–9147, 2018. 2