



OPEN

Spatio-temporal feature learning with reservoir computing for T-cell segmentation in live-cell Ca²⁺ fluorescence microscopy

Fatemeh Hadaeghi^{1✉}, Björn-Philipp Diercks³, Daniel Schetelig^{1,2}, Fabrizio Damicelli¹, Insa M. A. Wolf³ & René Werner^{1,2}

Advances in high-resolution live-cell Ca²⁺ imaging enabled subcellular localization of early Ca²⁺ signaling events in T-cells and paved the way to investigate the interplay between receptors and potential target channels in Ca²⁺ release events. The huge amount of acquired data requires efficient, ideally automated image processing pipelines, with cell localization/segmentation as central tasks. Automated segmentation in live-cell cytosolic Ca²⁺ imaging data is, however, challenging due to temporal image intensity fluctuations, low signal-to-noise ratio, and photo-bleaching. Here, we propose a reservoir computing (RC) framework for efficient and temporally consistent segmentation. Experiments were conducted with Jurkat T-cells and anti-CD3 coated beads used for T-cell activation. We compared the RC performance with a standard U-Net and a convolutional long short-term memory (LSTM) model. The RC-based models (1) perform on par in terms of segmentation accuracy with the deep learning models for cell-only segmentation, but show improved temporal segmentation consistency compared to the U-Net; (2) outperform the U-Net for two-emission wavelengths image segmentation and differentiation of T-cells and beads; and (3) perform on par with the convolutional LSTM for single-emission wavelength T-cell/bead segmentation and differentiation. In turn, RC models contain only a fraction of the parameters of the baseline models and reduce the training time considerably.

Regulation of cytosolic and organelle Ca²⁺ concentration and initial transient highly localized Ca²⁺ signals (Ca²⁺ microdomains) are essential for T-cell activation and initiation of effective immune responses^{1–5}. While mechanistic details of the initial intra-cellular Ca²⁺ elevation and propagation of Ca²⁺ microdomains during T-cell activation remain poorly understood, advances in fluorescence microscopy enabled monitoring subcellular structures and early signaling events throughout T-cell activation with finer spatial and temporal resolution^{1,4,6}.

With frame rates higher than 40 Hz⁴, a spatial resolution in the order of the diffraction limit and finer, and acquisition periods of several seconds to minutes, in-depth analysis of the imaging data requires efficient, ideally automated post-processing pipelines⁷. A central pipeline building block in live-cell imaging and Ca²⁺ microdomain analysis is the localization and segmentation of cells. Automated cell segmentation in high-resolution live-cell Ca²⁺ imaging data is, however, challenging due to an intrinsically low signal-to-noise ratio, fast Ca²⁺ signaling-based intensity fluctuations, overlaid by intensity changes on longer time-scales, due to, e.g., T-cell activation and photo-bleaching. Depending on the experimental setup, the cells further exhibit motion and deformation⁸. Moreover, if antibody-coated beads are used to mimic cell-cell interaction and to activate the cells^{4,9}, new objects with potentially similar intensity values and appearance than the cells enter the scene.

In this context, the present work describes computationally efficient segmentation approaches tailored to the requirements of live-cell Ca²⁺ microscopy and Ca²⁺ signaling analysis in T-cells. Methodically, the algorithms rely on the principles of reservoir computing (RC)¹⁰, which builds on the idea of recurrent neural networks

¹Department of Computational Neuroscience, Center for Experimental Medicine, University Medical Center Hamburg-Eppendorf (UKE), Martinistrasse 52, 20246 Hamburg, Germany. ²Center for Biomedical Artificial Intelligence (bAlome), University Medical Center Hamburg-Eppendorf (UKE), Martinistrasse 52, 20246 Hamburg, Germany. ³Department of Biochemistry and Molecular Cell Biology, Center for Experimental Medicine, University Medical Center Hamburg-Eppendorf (UKE), Martinistrasse 52, 20246 Hamburg, Germany. ✉email: f.hadaeghi@uke.de

(RNNs) to extract spatio-temporal features to achieve temporally consistent data analysis results and provides a computationally efficient model training and light-weight models in comparison to deep learning-based RNNs.

Related work. Image post-processing workflows for Ca^{2+} microscopy data commonly provide (semi-) automatic solutions to problems directly associated with the imaging process, such as bleaching correction, deconvolution, and emission-channel alignment in dual-wavelength measurements^{4, 7, 11–13}. Cell segmentation is often beyond the scope of standard toolkits; furthermore, due to the described peculiarities, existing solutions usually require higher levels of user involvement or suffer from limited generalizability. Besides, these techniques are highly susceptible to cell movement and deformation⁸. Furthermore, proposed combinations of traditional image processing methods^{14–18} are usually applied on a frame-by-frame basis. This bears a high risk of temporally inconsistent segmentation results for different frames in the presence of temporal intensity changes. In addition, limited generalization capability and prohibitive computational complexity pose problems for the segmentation of typically larger volumes of data recorded from different sets of cells and under different acquisition conditions.

The urge to develop generic segmentation and cell tracking algorithms, therefore, prompted the use of machine learning principles and is at the moment shaped by deep learning methods^{19, 20}. Currently dominating neural networks, however, usually comprise a feed-forward architecture, trained on static and independent frames. Due to the associated risk of temporally inconsistent segmentation results, we suggest utilizing recurrent neural networks (RNNs) to take advantage of temporal correlations in the data.

The applicability of recurrent neural networks, and especially deep learning-based RNNs, is nevertheless, at the moment, limited by the difficult and computationally expensive training process, making use of techniques such as temporal back-propagation²¹. Reservoir computing (RC) provides a computationally efficient alternative framework for RNN training^{22, 23}. Its properties make it interesting for the biomedical domain^{24, 25}, but, so far, applications are predominantly described for other fields^{22, 26, 27}, mainly focusing on signal processing tasks. In turn, RC application to (biomedical) image processing can only rarely be found^{28, 29}; and we are not aware of previous work on RC-based processing of spatio-temporal image data.

Contributions. Building on our previous work³⁰, we present, to the best of our knowledge, the first study that explores the capabilities of reservoir computing in the context of segmentation of spatio-temporal image series. Specifically, we developed RC algorithms that are suitable for application to single- and dual-wavelength Ca^{2+} imaging data and T-cell segmentation. The RC-based algorithms are compared to state-of-the-art deep learning architectures: a standard U-Net³¹ as de-facto standard in image segmentation and an U-Net-based convolutional long short term memory (LSTM)³² as problem-tailored state-of-the-art deep learning RNN solution.

Materials and methods

Reservoir computing for spatio-temporal image segmentation. The core to a reservoir computing model is a random, sparse, but fixed recurrent neural network, known as the *reservoir* (Fig. 1A), that non-linearly maps a time-dependent input signal into a higher dimensional signal space through the internal states of this dynamical system. The time-dependent output is computed as a linear combination of these reservoir variables. In contrast to traditional *deep* RNN training methods, RC only adapts the output weights to minimize an error measure (usually the mean squared error) between the desired target and the output signals. Thus, the neuron connections remain fixed, except for those from the reservoir toward the output layer, the so-called *readout connections*²¹. The non-linear expansion of the input signal into the high(er)-dimensional reservoir space, plus ease of training, enable RC models to efficiently learn to extract spatio-temporal features from time-dependent signals.

Using a general notation, the RC dynamics are governed by

$$\mathbf{x}(t + \Delta t) = f(\mathbf{W}\mathbf{x}(t) + \mathbf{W}^{in}\mathbf{u}(t + \Delta t)), \quad (1)$$

with $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ denoting the time-dependent N_x -dimensional reservoir state (i.e., a reservoir with N_x units), $\Delta t \in \mathbb{R}^+$ the time sampling period, $\mathbf{W} \in \mathbb{R}^{N_x \times N_x}$ and $\mathbf{W}^{in} \in \mathbb{R}^{N_x \times N_u}$ as internal and input weight matrices, respectively, and $\mathbf{u}(t) \in \mathbb{R}^{N_u}$ the input at time t . The internal states are updated via the non-linear function f .

The output $\mathbf{y}(t) \in \mathbb{R}^{N_y}$ is obtained from the extended system state $\mathbf{z}(t) = [\mathbf{x}(t); \mathbf{u}(t)]$ with $[\cdot; \cdot]$ as vertical vector concatenation by

$$\mathbf{y}(t) = g(\mathbf{W}^{out}\mathbf{z}(t)), \quad (2)$$

with g as an output activation function and $\mathbf{W}^{out} \in \mathbb{R}^{N_y \times (N_x + N_u)}$ the readout weight matrix. Training the RC system then means training the readout weights (depicted by dashed lines in Fig. 1A) by computing the linear regression weights of the target outputs on the already harvested states of the reservoir units and the inputs via ridge-regression¹⁰.

Encoding temporal image series into reservoir computing input data. A reservoir computing model in its standard formulation (i.e., Eq. 1) expects a single or multiple parallel time-series in the input. For temporal image processing applications, therefore, the temporal image series $(\mathbf{I}_i)_{i=1, \dots, N}$, $\mathbf{I}_i \in \mathbb{R}^{n_1 \times n_2}$ with n_1 and n_2 as number of pixels along the image axes must be converted into corresponding RC input data \mathbf{u} . In this study, we defined the following *encoding* schemes.

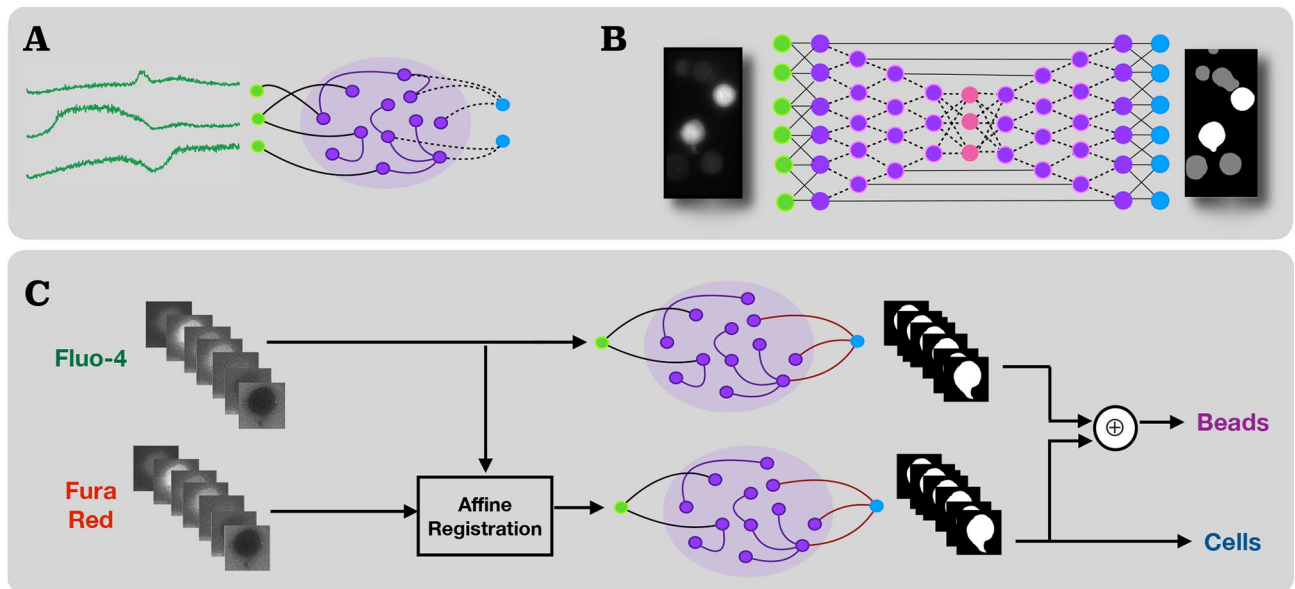


Figure 1. Illustration of the proposed models to segment and classify T-cells/beads in dual-wavelength Ca^{2+} imaging data. **(A)** Schematic representation of a reservoir computing model with fixed recurrent weights (solid lines) and trainable readout weights (dashed lines) commonly used for time-series processing. In the present work, temporal image series are converted into and processed as multiple parallel time-series (shown in green) via the encoding schemes detailed in the main text. **(B)** Schematic of a deep convolutional, U-Net-like neural network (dashed lines: trainable weights). **(C)** The proposed system for T-cell/bead segmentation and differentiation (Task 2) [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

Encoding scheme 1. Encoding scheme 1 is based on a straightforward vectorization of the images. For each image \mathbf{I}_i , six vectors $\mathbf{i}^{(k)} \in \mathbb{R}^{n_1 n_2}$ ($k = 1, \dots, 6$) were generated by $\mathbf{i}^{(1)} = \text{vec}(\mathbf{I}_i)$, $\mathbf{i}^{(2)} = \text{vec}(\mathbf{I}_i^T)$, $\mathbf{i}^{(3)}$ and $\mathbf{i}^{(4)}$ as forward- and $\mathbf{i}^{(5)}$ and $\mathbf{i}^{(6)}$ as backward-shifted versions of $\mathbf{i}^{(1)}$ and $\mathbf{i}^{(2)}$, i.e. $\mathbf{i}^{(3)} = [0, i_1^{(1)}, \dots, i_{n_1 n_2 - 1}^{(1)}]^T$ and $\mathbf{i}^{(5)} = [i_2^{(1)}, \dots, i_{n_1 n_2}^{(1)}, 0]^T$ and similar for $\mathbf{i}^{(4)}$ and $\mathbf{i}^{(6)}$. Forward- and backward-shifting as well as vectorization of the transposed image matrix aimed at providing spatial context to the reservoir. For a temporal image sequence $(\mathbf{I}_i)_{i=1, \dots, N}$, the input to the reservoir is eventually a real-valued matrix of size $7 \times (N n_1 n_2)$ with the first six rows corresponding to the $\mathbf{i}^{(k)}$ vectors for all N time points, and the seventh row containing a fixed bias. Thus, for the defined encoding scheme, the variable t defined in Eq. (1) does *not* directly refer to the temporal index of the image series frames, but to the sequential pixel order in the vectors. The reservoir state update still follows Eq. (1) with $\mathbf{u}(t)$ and $\mathbf{u}(t + \Delta t)$ denoting the t^{th} and $(t + 1)^{\text{th}}$ entities of the 7-dimensional input to the RC model (i.e., $\Delta t = 1$).

Using this encoding scheme, the segmentation task is formalized as a supervised binary classification problem with a single output node. The output node computes a linear combination of the reservoir states and returns a real-valued vector of length $N n_1 n_2$. A threshold function is then applied to map this vector to a $\{0, 1\}$ -vector of the same length; the threshold represents an additional hyperparameter. The resulting binary vector is finally re-ordered into the desired binary image series of length N .

Encoding scheme 2. The encoding scheme 2 is illustrated in Fig. 2. In comparison to the straightforward encoding scheme 1, it focuses on a pixel-level analysis and aims at a denser integration of spatial information. In detail, the number of input neurons of the reservoir is chosen to be $N_u = 9$, covering the intensity information of a 3×3 pixel neighborhood of a pixel for each time point of an image series of length N . Different to encoding scheme 1, in this case, variable t of Eq. (1) indeed refers to the temporal index, i.e. the frame number, of the considered image time series; $\mathbf{u}(t)$ and $\mathbf{u}(t + \Delta t)$ denote to the pixel intensities in the 3×3 neighborhood of the processed pixel at t and $t + 1$. Applied in a three-class segmentation context (see “Task 3: T-cell/bead segmentation and classification in single-emission measurements”), the encoding scheme is applied together with a RC model with a three-neuron output layer, returning the class-specific RC outputs for the considered pixel at a specific time point of the time series. RC inference for all image pixels leads to three temporal image series with N frames, which are converted into probability values via softmax layers. Based on the class-specific softmax values, eventually, N ternary images are generated.

Image acquisition and data characteristics. The image data in this study were acquired by live-cell fluorescence microscopy as detailed by Diercks et al.³³

Briefly, imaging was carried out with a Leica IRBE2 microscope (100-fold magnification) using a Sutter DG-4 as a light source at the image acquisition frequency of 40 Hz (data acquisition with Hamamatsu C9100 EMCCD

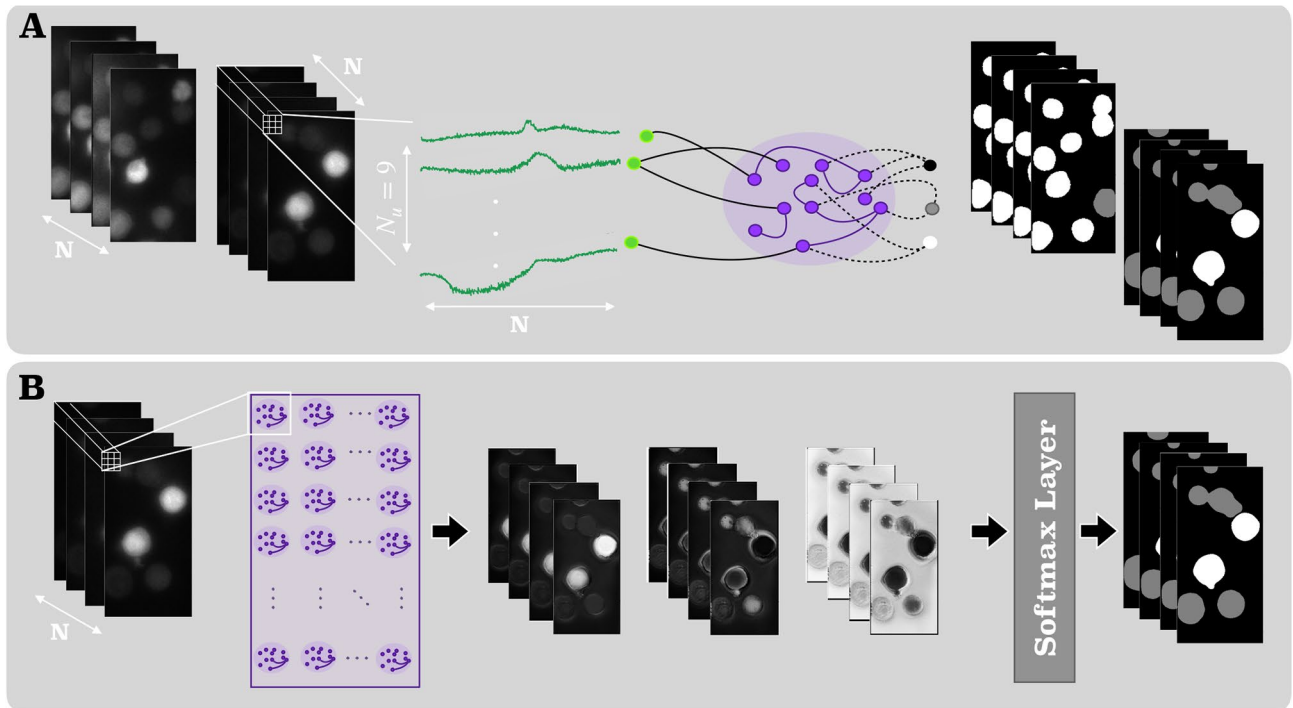


Figure 2. Illustration of the distributed RC structure for segmentation/classification (task 3) using encoding scheme 2. **(A)** Schematic of a single reservoir computing model trained on multitudes of $N_u = 9$ univariate time-series recorded at the sites of randomly sampled adjacent pixels. The three outputs are associated with ternary output classes. **(B)** Illustration of the system functioning during model inference [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

camera). A dual-view module (Optical Insights, PerkinElmer Inc.) was used to split the emission wavelengths of the two imaged Ca^{2+} indicators⁴.

Our experiments focused on Jurkat T-cells that typically exhibit significant motion and deformation during imaging, allowing us to better illustrate the advantages of the proposed segmentation algorithms. In addition, primary T-cell data were used to analyze generalizability capabilities of RC-based cell segmentation models. Primary T-cells are typically smaller and, from that perspective, more challenging to segment than Jurkat T-cells. They, however, exhibit less motion and deformation than Jurkat T-cells and are, therefore, less suited for analysis and illustration of the impact of integration of temporal information into the segmentation process.

All cells were loaded with Fluo-4 and Fura Red as cytosolic Ca^{2+} indicators and stimulated by beads coated with CD3-antibodies. The beads were added after several seconds of image acquisition. A typical image frame had a size of 500×250 pixel with a spatial resolution of 368 nm for each emission-wavelength; the resulting temporal sequence comprised > 7000 frames. Example data are shown in Fig. 3.

Application scenarios and experiments. We focused on three segmentation scenarios: (1) single object segmentation; (2) T-cell and bead segmentation and differentiation exploiting two emission-wavelengths information; and (3) T-cell/bead segmentation and differentiation in single emission-wavelength recordings. In the following, the developed RC algorithms are detailed and corresponding experiments described. The experiments were based on ten live-cell imaging recordings (computer hardware: Intel Xeon(R) E-2186 (3.80 GHz), 32 GB RAM, NVIDIA GeForce RTX 2080 Ti).

Task 1: single object segmentation. The first task aimed at illustrating general feasibility and an initial evaluation of RC-based object segmentation in spatio-temporal microscopy data. Given manually extracted regions of interest (ROIs) that include a single object and a set of sequential frames of the object, generic RC models were implemented to segment the object. Segmentation was performed on single emission-wavelength data, i.e., either Fluo-4 or Fura Red imaging data.

Reference segmentation (ground truth, GT) data for evaluation purposes was generated semi-manually. An unsupervised RC-based clustering model was trained for cell-customized pixel-wise data annotation. The model suggestions were visually presented to a human observer and rated as “well-labeled” or “bad”, reducing the laborious manual pixel-wise labeling to a binary classification task. The GT generation process (presented in Fig. S1) is detailed in Supplementary Note 1. 2000 ROIs with Jurkat T-cells (40 cells with 50 frames each; ROI size: 128×128 pixel) that were judged “well-labeled” were used for subsequent model training and evaluation (Fig. S2 illustrates samples of frames marked as successfully labeled). A subset of 280 frames (only Fluo-4 emission) was also manually re-labeled. The manually annotated data were used to investigate whether the GT

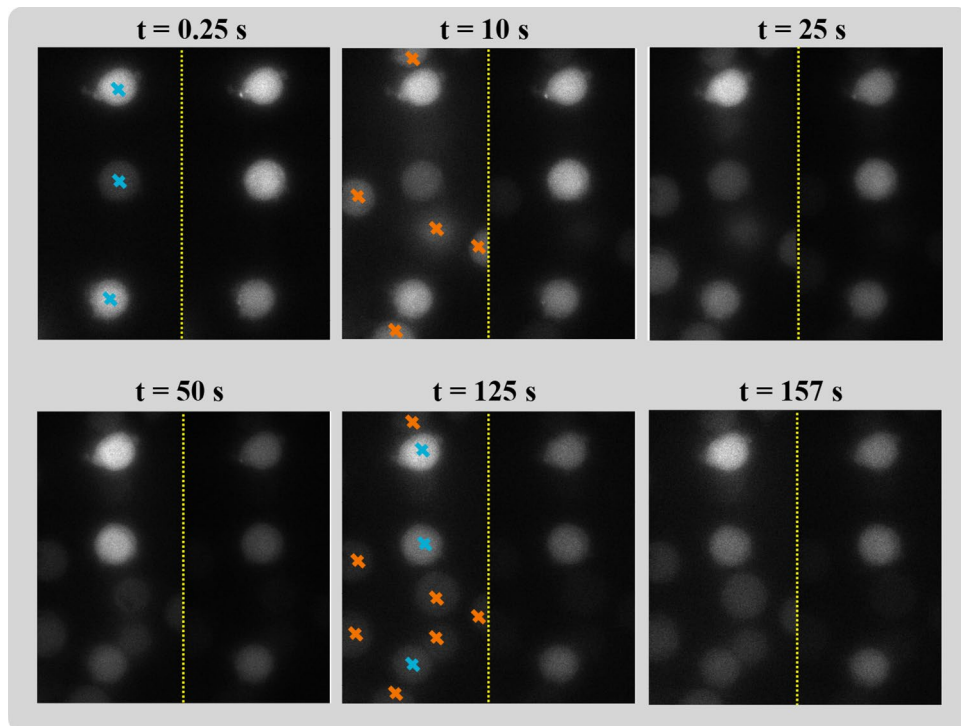


Figure 3. Six snapshots (with 500×500 pixels) of fluorescence intensity measured with Fluo-4 (left to the yellow border) and Fura Red (right) emissions. Cells are marked with blue crosses and beads with orange crosses at first appearance and, for further clarification, at $t = 125$ s. While there is no bead at $t = 2.5$ s, at $t = 10$ s, five and at $t = 25$ s, six beads are observable in the Fluo-4 images. The beads are not captured by Fura Red. Visible in all Fura Red images, three Jurkat T-cells with different sizes, shapes, and temporal activities are present [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

generation process resulted in a potential bias toward overestimation of RC segmentation accuracy. Finally, a set of 22 primary T-cells (50 frames each; same ROI size than for the Jurkat T-cells) were used to test whether the trained RC model is able to deal with imaging data for a different cell type not seen during training.

The RC hyperparameters were optimized by 5-fold cross validation using 500 of the 2000 ROIs (i.e., 10 cells; final parameters: $N_x = 100$, each neuron randomly connected to 10 neurons; activation function: tanh) based on encoding scheme 1. After parameter selection, the final RC model was trained on the entire 500 ROIs.

RC segmentation performance was compared to the U-Net (Fig. 1B; here: with a ResNet34-pretrained encoder³⁴) and a state-of-the-art deep learning RNN architecture for cell segmentation that integrates the idea of convolutional long short-term memory networks (C-LSTM) into the U-Net by substituting the standard convolutional layers of the encoder with C-LSTM layers that offer recurrent connections³². Training of the deep learning systems (hyperparameters were the default parameters) was performed on the same 500 ROIs used for RC training. As an additional classical baseline, we also applied Otsu thresholding.

The segmentation approaches were evaluated in detail on the 1500 ROIs and 30 Jurkat T-cells that were not used during training. The generalization capability of the trained RC model was further investigated on the above-mentioned 22 primary T-cells and the corresponding 1100 frames. Segmentation accuracy measures were pixel-wise accuracy and the Dice coefficient³⁵.

To also investigate the hypothesis that consideration of temporal correlations in the data as done in RNN-based models helps improving temporal consistency of segmentation results, a contour evolution analysis was performed. Therefore, the perimeter of the segmentation masks, its orientation (angle between the image x-axis and the major axis of the cell), and the mask area were evaluated for different models and the Jurkat T-cell data.

Task 2: T-cell/bead segmentation and differentiation using two-emission-wavelengths measurements. To demonstrate transferability of the task 1 results to a ‘real-world’ scenario, task 2 addresses the segmentation of full frames and temporal image data that contain multiple T-cells and antibody-coated beads. Thus, segmentation of T-cells not only means to reliably segment high intensity objects, but also to differentiate between T-cells and beads. Viewed in a single frame and emission wavelength (i.e., Fluo-4 or Fura Red), cells and beads can hardly be differentiated even by human observers (Fig. 3). The profound gradient between bead intensity values of corresponding Fluo-4 and Fura Red imaging data, however, can ameliorate the segmentation performance when the system is provided with the information of both cytosolic Ca^{2+} indicator emissions.

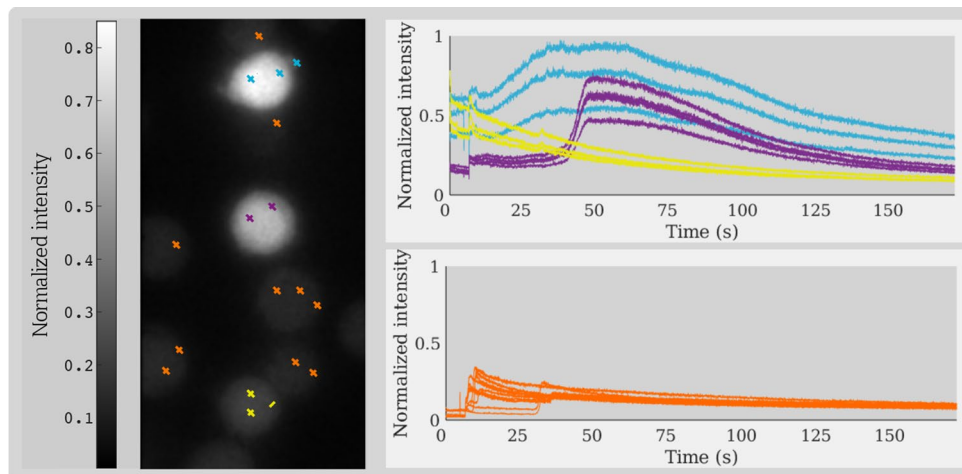


Figure 4. Examples of temporal signals recorded for Jurkat T-cell pixels (depicted in blue, purple and yellow) and antibody-coated beads (orange). The time-series highlight the notable difference in the fluorescence intensity levels obtained from different cells. Although the graphs recorded from sites corresponding to beads follow similar temporal patterns, there are lags between the times they appear in the imaging data. Near the end of the recording course, the fluorescence intensity at the bead sites are quite comparable to the cell intensities, making a frame-by-frame segmentation/classification ineffective [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

Therefore, suitable to be integrated into dual-wavelength Ca^{2+} imaging systems, we propose the RC-based segmentation and object classification scheme outlined in Fig. 1C. The system comprises two trained reservoir models: one reservoir directly receives Ca^{2+} images from Fluo-4 measurements, and the other one is presented with Fura Red sequences, affinely registered to the corresponding Fluo-4 frame to compensate for a potential misalignment of the different emission wavelength imaging information. Methodically, the two RC systems are identical to the approach described in “Task 1: single object segmentation”. Subsequent to object segmentation in both emissions, a logical XOR operation is applied to discriminate cells and beads.

Similar to task 1, a semi-manual RC-based annotation system (depicted in Fig. S3) was implemented to create ternary images (classes: background, T-cell, bead) and GT data as described in Supplementary Note 2. Example GT data are shown in Fig. S4. RC training was based on 1155 full-size frames (231 frames from 5 Jurkat T-cell image series, each frame with a size of 500×250 pixel; RC hyperparameters like in task 1, but $N_x = 500$). Testing was performed on a separate set of 1155 frames from five different Jurkat T-cell image series (frame number chosen due to RAM limitations).

The performance of the proposed RC algorithm was again compared to U-Net and U-Net-based LSTM results. To ensure comparability, the deep learning approaches were set up similar to the RC system: two models were trained, one using Fluo-4 and one using Fura Red information and the results combined via XOR. Training and test data were the same as used for the RC.

Task 3: T-cell/bead segmentation and classification in single-emission measurements. Simultaneous imaging of two Ca^{2+} indicators like Fluo-4 and Fura Red is motivated by the advantages of dual-wavelength ratiometric fluorescence microscopy: Computing the ratio between corresponding fluorescence intensity values allows, e.g., correcting for artifacts due to locally varying dye concentration, variations in laser intensity, and calculation of absolute Ca^{2+} concentrations³⁶. Furthermore, being able to use one excitation wavelength for two Ca^{2+} indicators has the advantage to detect local Ca^{2+} microdomains at a very high temporal and spatial resolution⁹.

However, aiming, for instance, at identification of players involved in the development of initial Ca^{2+} microdomains, a correlation of increased local cytosolic Ca^{2+} microdomains to cell organelles and Ca^{2+} channels is desirable, requiring staining the structures and measuring the corresponding fluorescence signal. For such scenarios, it is common to image the intracellular Ca^{2+} concentration using only a single Ca^{2+} indicator like Fluo-4. This, in turn, means that algorithms are required to differentiate T-cells and antibody-coated beads without extra information from other recording emission-wavelengths such as Fura Red in task 2.

To illustrate the complexity of this task, the temporal intensity profile of different cell and bead pixels are plotted in Fig. 4. The differentiation of cells and beads becomes almost impossible if taking into account the image information of only a single frame. The hypothesis is that RNNs are able to perform the task by making use of the distinct temporal patterns for bead and cell pixels.

To tackle the task by RC, we first re-used and evaluated the RC system and encoding scheme 1 as described for task 2. Further, to enforce the RC system to better preserve local spatial correlations of the data while simultaneously focusing on temporal pixel intensity patterns, we implemented and applied encoding scheme 2 (see “Encoding temporal image series into reservoir computing input data”).

	Input data characteristics		Reservoir parameters		
	Total number of frames per image series	Frame size	Number of neurons	Activation function f	Readout function g
Task 1	50 (each containing a single cell)	128×128	100	tanh	Identity
Task 2	231 (entire image frames)	500×250	100	tanh	Identity
Task 3	231 (entire image frames)	500×250	500	ReLU	Sigmoid

Table 1. Summary of the characteristics of the input data used in the present study and the chosen reservoir parameters for individual tasks.

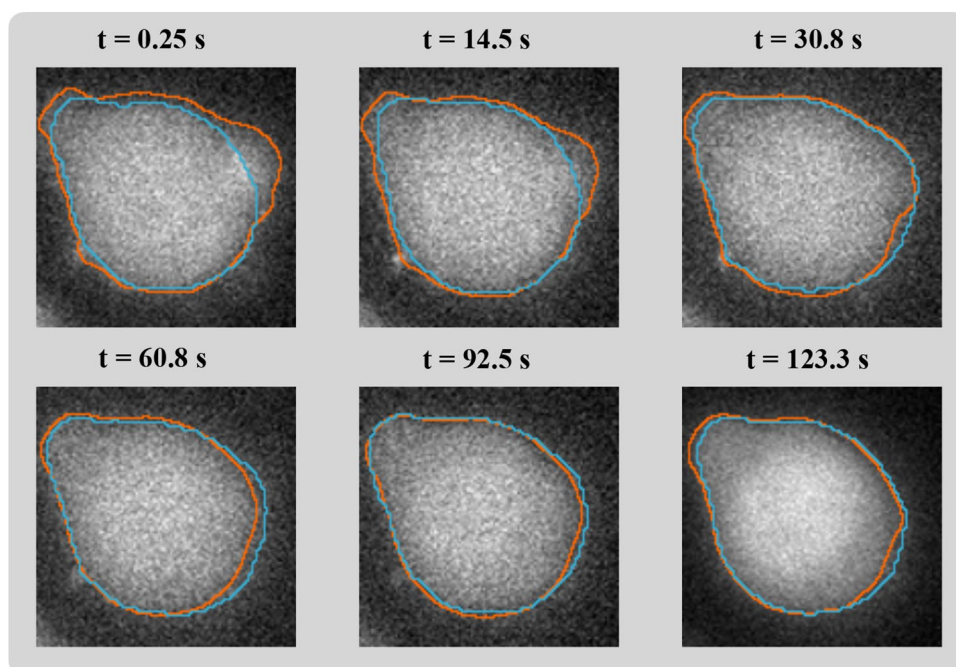


Figure 5. Illustration of task 1 results: Jurkat T-cell segmentation for six frames of a Fluo-4 Ca^{2+} imaging sequence. Blue: segmented mask border for the standard U-Net model. Orange: cell borders obtained by the reservoir computing model [Graphics created in MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

For the first RC system and encoding scheme, the training data was similar to the one used in task 2, except for using only the Fluo-4 imaging data. For the second RC encoding scheme, 20×10^7 pixels from the same data were selected for training. Hyperparameters were kept similar to the first encoding scheme, except for replacing the tanh activation function by ReLU. The test dataset was the same used for task 2.

The performance of the RC algorithms was compared to the results obtained by the standard U-Net and an adaptation of the U-Net-LSTM for multi-class classification.

Table 1 summarizes the RC input data characteristics and reservoir parameters for the individual tasks. For all tasks, the outputs of the different segmentation algorithms were post-processed following Arbellet al. ³² (i.e., application of morphological hole closing, removal of small segmented clusters) to avoid holes within the segmented objects and to reduce the number of false positive pixels. The post-processing parameters were identical for all segmentation approaches.

Results

Single object segmentation. Figure 5 shows segmentation results for an exemplary Jurkat T-cell and six frames for the RC algorithm and a standard U-Net. Both approaches achieve a fairly good segmentation quality. For some frames, the U-Net segmentation is closer to the visually perceived cell border, and for some, the RC results appear more appropriate. The same holds true for a comparison to the U-Net-LSTM. This visual impression is also reflected by the quantitative evaluation summarized in Table 2 (upper part): For the full test dataset, all machine learning-based segmentation approaches achieved accuracy values between 0.94 and 0.95 and Dice coefficients between 0.92 and 0.93 for both cytosolic Ca^{2+} indicator emissions (differences between algorithms or emissions not significant; testing by two-sample t tests; $p > 0.31$ for all comparisons). In comparison, the Dice values for Otsu thresholding, applied as a classical baseline approach, were between 0.89 and 0.91 ($p > 0.08$ for comparisons to the other segmentation approaches).

		Accuracy	Dice value
Evaluation using full semi-manual test set			
RC	Fluo-4	0.9453	0.9321
	Fura Red	0.9446	0.9315
U-Net	Fluo-4	0.9447	0.9297
	Fura Red	0.9415	0.9265
U-Net-LSTM	Fluo-4	0.9459	0.9321
	Fura Red	0.9369	0.9232
Otsu thresholding	Fluo-4	0.9402	0.8902
	Fura Red	0.9371	0.9096
Comparison of performance for manual/semi-manual subset			
RC	Manual	0.9242	0.9102
	Semi-manual	0.9252	0.9131
U-Net	Manual	0.9397	0.9233
	Semi-manual	0.9383	0.9232
U-Net-LSTM	Manual	0.8929	0.8748
	Semi-manual	0.8995	0.8862

Table 2. Quantitative evaluation of the segmentation accuracy of the considered machine learning models for **task 1**: reservoir computing, a standard U-Net, and an U-Net-based LSTM (evaluation metrics: mean class-wise accuracy and Dice coefficient). As a classical baseline approach, segmentation results for standard Otsu thresholding are also listed for the semi-manually generated test set. This test set consisted of 1500 frames captured from 30 Jurkat T-cells, i.e., 50 frames/cell. The subset used for the investigation of a potential bias due to the proposed semi-manual GT generation approach comprised a subset of 280 frames that were also manually labeled (Ca²⁺ indicator: Fluo-4).

The sub-analysis on the potential bias due to the GT generation process is summarized in Table 2 (lower part). There exist no significant difference between the metrics values for the semi- and the entirely manually annotated data that would indicate existence of a bias.

The results of the analysis of the temporal consistency of the segmentation masks are illustrated in Fig. 6 and supplemental video S1. The incorporation of temporal information by the RC system leads to smoother contour trajectories. While for the U-Net (i.e., frame-by-frame segmentation), the contour length, the mask orientation, and the mask area show abrupt changes between different frames, respective measures for the RC system show a smoother and more plausible evolution. For the deep learning-based RNN approach (U-Net-LSTM), the results are similar to the RC system (presented in Fig. S5, supplementary document).

Application of the RC model trained for segmentation of Jurkat T-cells to unseen primary T-cells led to a drop of the Dice values compared to Jurkat T-cell segmentation. For Fluo-4 emission measurements, the accuracy and the Dice value were 0.9476 and 0.8137, respectively; for Fura Red emission data, the accuracy and the Dice value were 0.9674 and 0.8670. The Dice values indicate a potential overfitting of the RC model to Jurkat T-cell data characteristics. However, Otsu thresholding applied to the same data yielded Dice values of 0.7189 (Fluo-4; $p = 0.004$ for comparison to RC Dice values) and 0.8138 (Fura Red; $p = 0.008$). Thus, the information learned by the RC model appears helpful compared to the basic intensity-based two-class pixel differentiation even for the different cell type data.

T-cell/Bead segmentation and differentiation using two-emission-wavelengths measurements. The results for task 2 are summarized in Table 3. Given the different appearance of the beads for the two emission wavelengths, this task appears relatively straight-forward. However, the accuracy and Dice values indicate that incorporation of temporal information already helps improving segmentation and classification performance for this task: While the RC and the U-Net-LSTM systems perform on par, they both outperform the standard U-Net that was applied frame-by-frame (statistical testing omitted due to limited number of independent samples, i.e., five imaging sequences). In addition, the segmentation of the beads appears to be more complex for all three systems, with the standard U-Net almost entirely failing.

T-cell/bead segmentation and classification using single-emission measurements. The quantitative results for task 3 are summarized in Table 4. The standard U-Net was, similar to task 2, not able to provide acceptable results and was discarded from further analyses.

Table 4 illustrates that the RC encoding style, i.e., the approach to convert the images into a format that is suitable for RC-based processing, plays an important role. Compared to the straightforward encoding scheme 1 that is based on direct vectorization of the image matrices, the second scheme led to a significant increase of segmentation accuracy. Exemplary RC (with encoding scheme 2) segmentation results are shown in Fig. 7 and the supplementary video S2. A separation of close-by beads is, however, not always feasible; this remains for further methodological refinement.

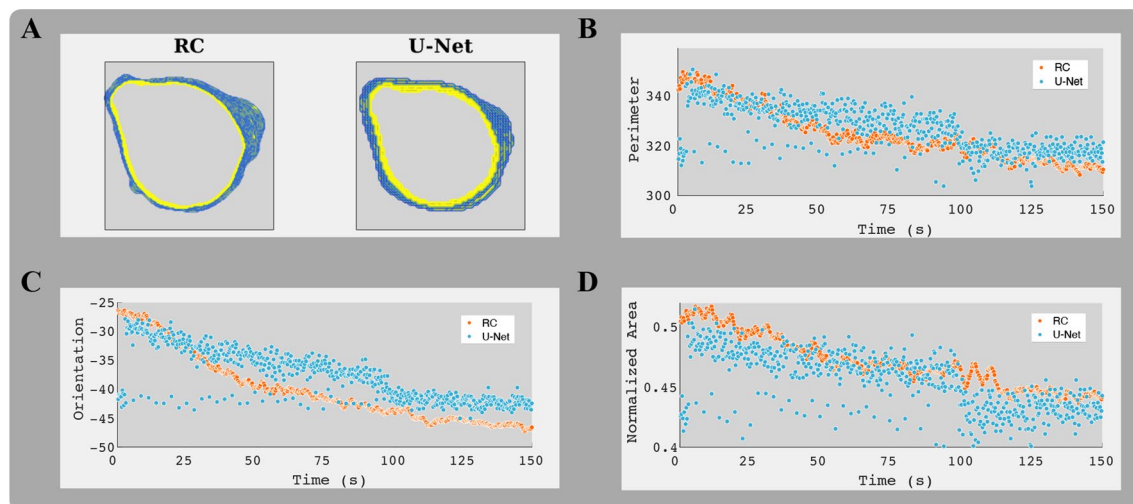


Figure 6. Predicted temporal contours evolution for a representative Jurkat T-cell and segmentation in the Fluo-4 emission data. **(A)** Unlike the discrete contour trajectory returned by the standard U-Net, the contours predicted by the RC method evolve continuously. **(B)** Frame-to-frame changes in contour length (in number of pixels) of the predicted masks. **(C)** Mask orientation, the angle (in degree) between the horizontal image axis (i.e., x-axis) and the major axis of an ellipse with the same second-moments as the segmented object, over time. **(D)** Normalized mask area over time. The area of the object was divided by the total number of pixels in each frame. The same illustration for the results of the U-Net-LSTM is presented in the supplementary materials (Fig. S5) [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

	Objects	Accuracy	Dice value
RC	T-cells	0.9653	0.8795
	Beads	0.9518	0.5819
U-Net	T-cells	0.9367	0.8164
	Beads	0.9437	0.1389
U-Net-LSTM	T-cells	0.9631	0.8842
	Beads	0.9483	0.5836

Table 3. Performance evaluation of the reservoir computing model and the deep learning models for **task 2** (test set based on 1155 frames extracted from five Ca^{2+} imaging sequences).

	Objects	Accuracy	Dice value
RC (encoding 1)	T-cells	0.9226	0.7623
	Beads	0.9266	0.3534
RC (encoding 2)	T-cells	0.9520	0.8528
	Beads	0.9459	0.5909
U-Net-LSTM	T-cells	0.9497	0.8540
	Beads	0.9579	0.6242

Table 4. Performance evaluation of reservoir computing models and the U-Net-based LSTM for Jurkat T-cell and bead segmentation and differentiation in single-emission recordings (**task 3**; test set: 1155 frames from five imaging sequences). Corresponding class-wise normalized confusion matrices are presented in Fig. S6.

Compared to the RC models, the mean accuracy and Dice values for the U-Net-LSTM are, although being in the same range, slightly higher. It should, however, be kept in mind that the implemented RC models have a drastically lower number of trainable parameters (approximately 1500 in the current study) than the U-Net-LSTM ($> 6.5 \times 10^6$) and the standard U-Net ($> 3.6 \times 10^9$). In the current experiment, this led to a reduction of training time from 26 h for the U-Net-LSTM to 1 h for the RC system, although the U-Net-LSTM training was performed on GPU and was already highly optimized for GPU usage, while the RC training was on CPU and was not optimized for parallel computing.

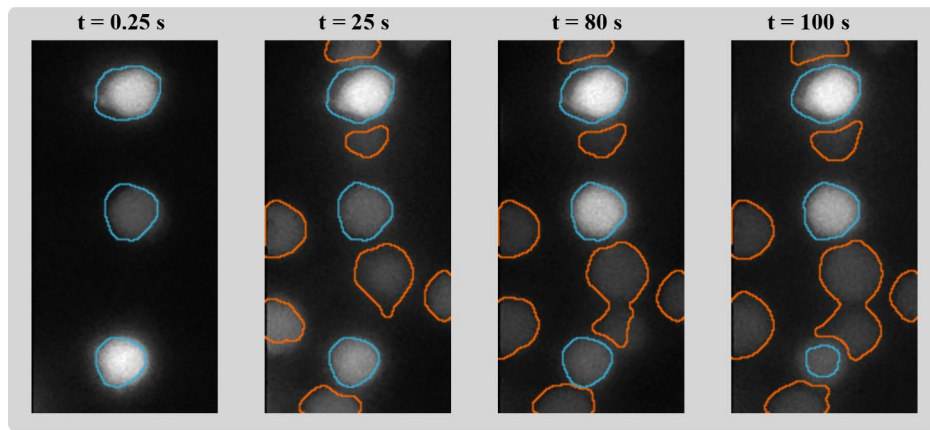


Figure 7. Boundaries of segmented Jurkat T-cells (blue) and anti-body coated beads (orange) in four frames of Fluo-4 Ca^{2+} imaging data, computed by the proposed RC model with encoding scheme 2. Despite cellular and beads movement and significant local and temporal variation of fluorescence intensity, the algorithm provides a robust and efficient solution to the task [Graphics created with MATLAB and the Image Processing Toolbox Release 2019b, The MathWorks, Inc. (<http://mathworks.com/products/image.html>)].

Discussion

High-resolution Ca^{2+} imaging methods allow characterizing spatio-temporal dynamics of initial Ca^{2+} signaling in T-cells, a fundamental process in the adaptive immune system. The increasing amount of acquired data results in a need for efficient image processing and analysis solutions. The present study explores the potential of reservoir computing (RC) for temporally consistent object and, in particular, T-cell segmentation in spatio-temporal Ca^{2+} imaging data. The underlying rationale was that RC represents a computationally efficient RNN-based approach to learn spatio-temporal features and can help to overcome drawbacks of current deep learning systems.

Applied to Jurkat T-cell segmentation as well as bead and cell segmentation and classification using either single- or two-emission wavelengths imaging information, the RC models perform in terms of segmentation accuracy at least on par with the de-facto standard in biomedical image segmentation, the standard U-Net. For differentiation of T-cells and beads, which requires integration of temporal information, RC outperforms the U-Net, demonstrating the potential of spatio-temporal learning inherent to the RC paradigm.

Compared to a U-Net-based LSTM as a state-of-the-art RNN architecture, the RC models show a similar performance both in terms of segmentation accuracy and temporal consistency of the segmentation results. At this, it should be noted that the semi-manual GT generation pursued in the present study included a frame-by-frame visual quality check and application of frame-specific thresholds. This partly led to temporal inconsistencies of the annotations for the temporal images series used for system training (see supplemental video S2). The obtained temporally mainly consistent segmentation results therefore also illustrate a certain degree of robustness of both RNN approaches with respect to corresponding training data imperfections. However, convolutional LSTMs are computationally expensive, difficult to train, and, consequently, still rarely applied in biomedical context. In the current work, the LSTM training took, for instance, more than a day on GPU, while RC training required one hour on CPU. Furthermore, the RC model comprised 1500 trainable parameters—whereas the LSTM $> 6.5 \times 10^6$ parameters. Thus, a similar segmentation performance was achieved with only 0.023% of trainable parameters.

Despite faster training, due to the current CPU implementation, the proposed RC-based image segmentation is, however, not real time-capable: RC inference for a single 128×128 pixel frame takes approximately 0.5 s for the described hardware, and the inference time scales with the number of pixels. Optimization for parallel computing and re-implementation for GPU usage is, nevertheless, expected to result in a significant shortening of RC inference times especially for the proposed encoding scheme 2, rendering real-time RC segmentation realistic.

With regard to the presented results, we would like to note that the spatio-temporal Jurkat T-cell image series considered in our study are representative for the imaging conditions and data characteristics at our laboratory^{4,9,33}. However, it remains to be shown that our methods and observations can be transferred to and confirmed for different, maybe larger or more heterogeneous datasets and data acquired under different imaging conditions. In particular altered imaging conditions, but also cell and cell dynamics characteristics not present in the training data will necessitate retraining the models. This becomes already evident by the drop of the Dice values seen for segmentation of primary T-cells by means of the RC model trained for Jurkat T-cell segmentation (see results for task 1). To foster testing of the proposed approaches on other datasets, we provide the RC source code, together with the models and example data, as open source (see Data Availability statement).

For future work from a method perspective, it remains to extend our RC architecture by additional reservoir layers to extract multiple-scale temporal and/or spatial features. We expect the incorporation of multi-scale information to further improve segmentation accuracy.

Conclusions

The current work demonstrates reservoir computing to be an efficient alternative to computationally expensive deep learning-based networks for temporally consistent cell segmentation in high-resolution live-cell Ca^{2+} imaging.

Data availability

The RC source code, the models, and example data are provided publicly available at https://github.com/IPMI-ICNS-UKE/Jurkat_cell_segmentation.

Received: 10 November 2020; Accepted: 25 March 2021

Published online: 15 April 2021

References

- Russell, J. T. Imaging calcium signals in vivo: A powerful tool in physiology and pharmacology. *Br. J. Pharmacol.* **163**, 1605–1625 (2011).
- Trebak, M. & Kinet, J. Calcium signalling in T cells. *Nat. Rev. Immunol.* **1**, 20 (2019).
- Wolf, I. & Guse, A. Ca^{2+} microdomains in t-lymphocytes. *Front. Oncol.* **7**, 73 (2017).
- Diercks, B.-P. *et al.* ORA11, STIM1/2, and RYR1 shape subsecond Ca^{2+} microdomains upon T cell activation. *Sci. Signal.* **11**, eaat0358 (2018).
- Diercks, B.-P. & Guse, A. H. Unexpected players for local calcium signals: STIM and ORAI proteins. *Curr. Opin. Physiol.* **20**, 20 (2020).
- Randriamampita, C. & Lellouch, A. Imaging early signaling events in T lymphocytes with fluorescent biosensors. *Biotechnol. J.* **9**, 203–212 (2014).
- Schetelig, D. *et al.* A modular framework for post-processing and analysis of fluorescence microscopy image sequences of subcellular calcium dynamics. In *Bildverarbeitung für die Medizin 2015* 401–406 (Springer, 2015).
- Antoni, S. *et al.* Systematic analysis of Jurkat T-cell deformation in fluorescence microscopy data. In *Bildverarbeitung für die Medizin 2017* 275–280 (Springer, 2017).
- Wolf, I. *et al.* Frontrunners of t cell activation: Initial, localized Ca^{2+} signals mediated by NAADP and the type 1 ryanodine receptor. *Sci. Signal.* **8**, ra102–ra102 (2015).
- Lukoševičius, M. & Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009).
- Hodgson, L., Nalbant, P., Shen, F. & Hahn, K. Imaging and photobleach correction of mero-cbd, sensor of endogenous Cdc42 activation. *Methods Enzymol.* **406**, 140–156 (2006).
- Giovannucci, A. *et al.* Caiman an open source tool for scalable calcium imaging data analysis. *Elife* **8**, e38173 (2019).
- Wallis, J. W., Miller, T. R., Lerner, C. A. & Kleerup, E. C. Three-dimensional display in nuclear medicine. *IEEE Trans. Med. Imaging* **8**, 297–300 (1989).
- Fan, G., Zhang, J.-W., Wu, Y. & Gao, D.-F. Adaptive marker-based watershed segmentation approach for t cell fluorescence images. In *International Conference on Machine Learning and Cybernetics*, vol. 2, 877–883 (IEEE, 2013).
- Nordenfelt, P., Elliott, H. L. & Springer, T. A. Coordinated integrin activation by actin-dependent force during T-cell migration. *Nat. Commun.* **7**, 13119 (2016).
- Jiang, T., Yang, F., Fan, Y. & Evans, D. J. A parallel genetic algorithm for cell image segmentation. *Electron. Notes Theoret. Comput. Sci.* **46**, 214–224 (2001).
- Lee, A. M., Colin-York, H. & Fritzsche, M. Calquo 2: Automated Fourier-space, population-level quantification of global intracellular calcium responses. *Sci. Rep.* **7**, 1–11 (2017).
- Salles, A. *et al.* Barcoding t cell calcium response diversity with methods for automated and accurate analysis of cell signals (maacs). *PLoS Comput. Biol.* **9**, e1003245 (2013).
- Falk, T. *et al.* U-net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
- Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W. & Rusu, M. A deep learning-based algorithm for 2-d cell segmentation in microscopy images. *BMC Bioinform.* **19**, 1–11 (2018).
- Polydoros, A. S., Nalpanitidis, L. & Krüger, V. Advantages and limitations of reservoir computing on model learning for robot control. In *IROS Workshop on Machine Learning in Planning and Control of Robot Motion* (2015).
- Jaeger, H. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. *Bonn, Germany, German National Research Center for Information Technology (GMD) Technical Report*. (2001).
- Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
- Hadaeghi, F. Reservoir computing models for patient-adaptable ECG monitoring in wearable devices. [arXiv:1907.09504](https://arxiv.org/abs/1907.09504) (arXiv preprint) (2019).
- He, X., Liu, T., Hadaeghi, F. & Jaeger, H. Reservoir transfer on analog neuromorphic hardware. In *9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 1234–1238 (IEEE, 2019).
- Jaeger, H. & Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
- Triefenbach, F., Jalalvand, A., Schrauwen, B. & Martens, J.-P. Phoneme recognition with large hierarchical reservoirs. *Adv. Neural Inf. Process. Syst.* **20**, 2307–2315 (2010).
- Meftah, B., Lezoray, O. & Benyettou, A. Novel approach using echo state networks for microscopic cellular image segmentation. *Cogn. Comput.* **8**, 237–245 (2016).
- Souahlia, A. *et al.* Echo state network-based feature extraction for efficient color image segmentation. *Concurr. Comput. Pract. Exp.* **20**, e5719 (2020).
- Hadaeghi, F., Diercks, B.-P., Wolf, I. M. & Werner, R. Reservoir computing for Jurkat T-cell segmentation in high resolution live cell Ca^{2+} fluorescence microscopy. In *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 1587–1591 (IEEE, 2020).
- O. Ronneberger, P. F. & Brox, T. u-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (Springer, 2015).
- Arbelle, A. & Raviv, T. R. Microscopy cell segmentation via convolutional lstm networks. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 1008–1012 (IEEE, 2019).
- Diercks, B.-P., Werner, R., Schetelig, D., Wolf, I. M. A. & Guse, A. H. High-Resolution Calcium Imaging Method for Local Calcium Signaling. In *Calcium-binding proteins of the EF-hand superfamily* Vol. 1929 (ed. Heizmann, C. W.) 27–39 (Springer, 2019).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778, (2016).
- Dice, L. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945).

36. Bootman, M., Niggli, E., Berridge, M. & Lipp, P. Imaging the hierarchical Ca^{2+} signalling system in HeLa cells. *J. Physiol.* **499**, 307–314 (1997).

Acknowledgements

We would like to thank Assaf Arbelle for helpful conversations and feedback on implementation of the U-Net-LSTM models.

Author contributions

F.H. is the corresponding author. She carried out theoretical studies, data analysis, co-wrote and co-edited the paper. B.P.D. carried out experimental studies, constructed the experimental set up and co-edited the paper. D.S. and F.D. carried out data analysis, actively participated in discussions of the data observed and co-edited the paper. I.M.A.W. designed and supervised experimental studies. R.W. supervised theoretical studies, co-wrote and co-edited the paper. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) (project number 335447717; SFB1328, project A02 to B.-P.D. and R.W.).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87607-y>.

Correspondence and requests for materials should be addressed to F.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021