SC-Track: a robust cell tracking algorithm for generating accurate single cell linages from diverse cell segmentations

Chengxin Li[1,2], Shuang Shuang Xie[2], Jiaqi Wang[1,2], Septavera Sharvia[4], Kuan Yoow Chan[1,2,3]

[1] Department of Cardiovascular Medicine, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310058, P. R. China.

[2] Centre for Cellular Biology and Signalling, Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining, 314400, P. R. China.

[3] College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, EH4 2XR, UK.

[4] Department of Computer Science, University of Hull, Hull, HU6 7RX, UK.

Correspondence to:

kychan@intl.zju.edu.cn; kuanyoow.chan@outlook.com

Abstract

Computational analysis of fluorescent timelapse microscopy images is a powerful approach to study biological processes in detail. Core to this approach is the generation of accurate single cell linages from cell segmentations for reliable quantitative analysis. Convolutional neural networks (CNNs) are increasingly being used to segment and classify cells in microscopy images, but current cell tracking solutions are sensitive to inaccurate cell segmentations from CNNs. We present SC-Track, a cell tracking algorithm that employs a hierarchical probabilistic cache-cascade model. Our results show that SC-Track generates accurate single cell linages without parameter tuning, from cell segmentations of varying qualities, morphological appearances, and imaging conditions. Furthermore, SC-Track is equipped with a cell class correction feature to improve the accuracy of multi-class cell classifications in a time series. These features make SC-Track a robust generalist cell tracking algorithm that works with diverse segmentation outputs from CNNs to generate accurate cell linages and classifications.

Keywords: timelapse microscopy imaging, single cell tracking, cell division, deep learning, convolutional neural networks, cell cycle.

Main text

The analysis of time resolved fluorescent microscopy images to obtain cellular dynamics at the single cell level has enabled the detailed study of intracellular signalling events previously invisible to conventional cell biological approaches[1,2]. This method has led to the delineation of key signalling pathways that induce a variety of cell fate decisions[3–7]. Core to these approaches is the use of fluorescent markers to mark single cells, quantify signalling events and classify cellular states. The generation of single cell tracks from these fluorescent timelapse microscopy images is often a challenging process, requiring extensive optimisations of fluorescent markers and imaging conditions. This is to ensure that optimal cell segmentations are obtained as they are essential for accurate single linage tracing and reliable mother-daughter assignments[2]. To generate good quality fluorescent images, the prolonged exposure of live cells to high intensity ultraviolet light is a major limitation. This is because excessive

phototoxicity from exposure from ultraviolet light can result in cellular stress or death, making this approach impractical as an approach to study long term biological events[1].

To bridge this limitations, deep learning based convolutional neural networks (CNNs), have been employed in a variety of approaches to overcome the inherent limitations of conventional fluorescence-based microscopy approaches[8]. Among the most successful applications are the use of autoencoder CNNs, enabling computationally efficient image restoration of microscopy images for deconvolution, denoising and generating super-resolution image reconstructions[9]. Another area where CNNs have been successfully deployed is in the automated segmentation and classification of microscopy images[10–13]. CNNs have been demonstrated to perform very well in automatically detecting, segmenting and classifying heterogenous cellular features of microscopy images, a task that often requires time consuming manual human annotations[13–16].

However, the application of deep learning CNNs in the automated segmentation and classification of fluorescent microscopy images presents another challenge for reliable cell tracking. This is caused by the stochastic nature of the cell segmentations derived from these deep learning-based image analysis approaches[17]. Under ideal conditions, state-of-the-art deep learning approaches such as Mask RCNN, U-Net, Cellpose and StarDist often fail to accurately detect and classify all objects instances[12,13,17,18]. Thus, it is generally accepted that the segmented images from deep learning methods will be inherently noisy with instances where objects fail to be detected or are misclassified. These inaccuracies pose a major challenge for widely used cell tracking approaches to generate accurate single cell tracks, limiting the utility of these deep learning methods.

To overcome this inherent limitation, we developed a novel cell tracking algorithm called Single Cell Track (SC-Track). It employs a hierarchical probabilistic cache-cascade model to overcome the noisy output of deep learning models (Fig. 1). We show that SC-Track can generate robust single cell tracks from noisy segmented cell outputs ranging from missing segmentations and false detections. In addition, SC-Track can take noisy cell instance classifications and provide smoothed classification tracks to aid the accurate quantification and classification of cellular events. Finally, SC-Track has a built-in biologically inspired cell division algorithm that can robustly assign mother-daughter

associations from segmented nuclear or cellular masks, enabling high fidelity single cell tracking of cellular events over multiple cell generations.

Results

Tracking algorithm overview

SC-Track employs a tracking-by-detection approach, whereby detected cells are associated between frames. A TrackTree data structure was used (Fig. 1), to store the tracking relationships between each segmented cell temporally and spatially. Each branch of the TrackTree represents a single-cell linage of the tracked instance of a segmented cell, where branch divisions indicate cell division events, and the nodes on the branches represent the segmented instances of individual cells in a specific frame. Contained in each node of the TrackTree branch are the extracted features of the segmented cell.

During the tracking process, SC-Track initializes the TrackTree list with all cells from the initial frame, representing the initial single-cell tracks for the entire time-lapse sequence. To reduce computational costs, SC-Track will attempt to connect each segmented instance with its corresponding cell from the previous frame using a hierarchical tracking approach. SC-Track will initially examine the intersection over union (IoU) of the area between segmented cells between the current frame and preceding frame (Fig. 2). Segmented cells with only one overlapping segmentation are assumed to a high confidence linked cell and is automatically assigned to the corresponding TrackTree. In situations where there are multiple segmented cells with overlapping IoUs, SC-Track will assign segmented cells by maximising the similarity index between candidate segmented cells between frames. If there are no segmented cell in the current frame overlapping with a segmented cell from the previous frame, SC-Track will expand the search area (default = 1), to identify possible tracking candidates.

Using this method of recursive searching of candidate segmented cells from the previous frame, virtually all segmented cells can be accurately assigned to the correct TrackTree. In the event where there are more segmented cells than the number of cached TrackTrees, three possible scenarios will be considered: (1) The orphan segmented cell is a false detection; (2) The segmented cell is a true detection that recently entered into the field of view due to cell migration; (3) A cell division event has occurred

leading to the generation of 2 or more (in the case of multi-polar mitosis) daughter cells. The approach used to resolve these possible scenarios will be discussed in the subsequent two sections.

Detecting and assigning cell division events

When there are more segmented cells than the number of cached TrackTrees, SC-Track will determine if a cell division event has occurred (Fig. 3). A cell division event is deemed to have occurred when SC-Track is able to match a mother cell from the previous frame to the daughter cells in the subsequent frame. To achieve this, SC-Track will first determine if there are putative mother cells in the mitotic state in the previous frame. If the segmented cell contains cell cycle classifications, SC-Track will allow cell division events to occur at the TrackTree nodes where the mother cell is classified to be in mitosis (M phase). However, if no cell cycle information is available, this process is not enabled, and SC-Track will attempt to determine if a cell division event has occurred by matching orphan segmented cells to a potential mother cell using a cell cycle independent approach.

To enable robust detection of cell division events in the absence of cell cycle data, SC-Track applies a series of rules based on well-established principles observed from mammalian cells undergoing cell division[19,20]. When assigning a potential mother-daughter association from a potential cell division event, the following criteria must be met: (1) At least one candidate cell that cannot be accurately matched to other cells were found; (2) The segmented mother cell in the previous frame must be at least 1.3 × the size of the segmented daughter cells in the following frame. (3) The candidate mother cell that has not undergone a cell division event recently (20 frames by default). (4) A candidate mother cell is identified in the expanded search area of the unlinked segmented cell. If a suitable candidate mother cell is found in the previous frame for the orphan segmented cell, the TrackTree will be branched accordingly. However, if no suitable candidate mother cell was found, SC-Track will assume that this is a new detection event and assign a new TrackTree to the segmented orphan cell.

Cache matching frames to address false and missing detection events

Due to the stochastic nature of CNNs in detecting cells, there is a possibility that true cell instances fail to be detected or false detections may arise[12,13,17,18]. To overcome the stochastic loss of true instances in

the segmented cells, a cache matching system was developed. In an event where a TrackTree is unable to find a matching cell in the current frame, it is assumed by default that this was caused by a failed detection event. The TrackTree will be cached for five consecutive frames. If a matched segmentation was found within five frames, SC-Track will automatically assign the matched segmented cell to the corresponding TrackTree and the intervening gaps automatically filled with segmentations from the cache memory from the last detected instance. In the event where no matching segmented cells was found in the next five frames, the specific branch of TrackTree will be inactivated and can no longer be used to track cells in subsequent frames. Short TrackTree initialisations (user defined, default = 10 nodes), will be removed at the end of the tracking process to remove false detection instances.

Instance classification smoothing

Instance segmentation of cells from deep learning models that classify more than one class are often challenged with noisy classifications[21]. To address this, we have implemented a class smoothing function to smooth out noisy classification of cells that transition from one cellular state to another. We developed the TrackTree Class Smoothing (TCS) algorithm (Fig. 3) to automatically correct the predicted results of cell type classifications. TCS assumes that a cell classification change is more likely to be accurate in a time series when the same cell is classified with the same classification over several frames. To evaluate the accuracy of the cell class change, TCS adopts a probabilistic cached search model. This search process is confined to the individual branch of the TrackTree and does not extend beyond the cell division branch.

The TCS probabilistic cached search model functions with the following logic: During the initialisation of the TrackTree, TCS will automatically adopt the initial classification of the detected cell instance as the default class. When TCS detects an instance where the tracked cell undergoes a cell classification change to Type A, the algorithm will undertake a cached forward search on the TrackTree (default search window = 10 frames) to count the number of occasions the tracked cell is classified as Type A. If the number of nodes classified as Type A exceeds a probability threshold (default = 6), TCS will conclude that a change in cell classification has occurred and will update the default classification as Type A. Otherwise, the node where Type A was first detected will be assumed to be wrong and

corrected to the previous default cell classification. Exiting the default Type A classification occurs when a Type A' classification is detected, TCS performs a forward TrackTree cached search (default = 10 nodes) for the frequency for the Type A' classification. If the number of nodes of Type A' classifications exceed the probability threshold (default = 6), it is considered that the Type A classification has ended and the new default Type A' classification is adopted. This process can be repeated to multiple cell classifications.

SC-Track cell tracking performance evaluation

SC-Track overall cell tracking performance was measured using two metrics, the Multi-Object Tracking Accuracy (MOTA)[22,23] and the harmonic mean of Identification Precision and Recall (IDF1)[24]. We also introduced a new metric called the Cell Division F1 score (CDF1), to measure the cell tracker's ability to reliably detect cell division events and accurately assign mother-daughter cell relationships. For comparison, we benchmarked SC-Track against three other freely available cell tracking algorithms that provide similar functionalities: TrackMate[25,26], Deepcell-tracking[27], and pcnaDeep[28]. Initial tests focused on generating single cell tracks from nuclear masks obtained in ideal conditions, using manually corrected nuclear segmentation masks with accompanying cell cycle classifications with 5-minute temporal resolutions (Fig. 4). The results show that with ideal segmentation results, SC-Track gave the best performance, and the top three trackers gave a score > 0.9 in both metrics. We then assessed the performance of SC-Track in tracking cell division events by comparing the CDF1 score. SC-Track gave the best performance giving a CDF1 score of > 0.9 in all five test datasets (Fig. 4).

To further measure the reliability of SC-Track in generating accurate single cell linages, we resampled our original test dataset to mimic imaging time intervals of 10, 15 and 20 minutes. The increase in time intervals poses a more challenging cell tracking problem, as each cell in a field of view has more time to migrate spatially and the change in its cellular morphology between the preceding and subsequent frame will be larger. Our results show that SC-Track gives the best IDF1 scores in the 5-minute interval, but its performance is reduced at longer time intervals (Fig. 4). For the MOTA score, pcnaDeep maintained the best overall scores. These mixed results displayed by IDF1 and MOTA is caused by the differences in how each metric calculates tracking accuracy. IDF1 is more sensitive to the

total duration of incorrect track assignments while MOTA is more sensitive to the total number of track switches[24]. More importantly, SC-Track CDF1 scores were considerably better than the next best tracker pcnaDeep across all time intervals (Fig. 4). These results indicate that SC-Track gave the best overall performance and works well across varying temporal resolutions.

State-of-the-art deep learning-based CNN instance segmentations are generally known to display a low number of instance segmentation errors[12,13,17,18]. These range from missing segmentations to inaccurate segmentations, where cell instances are improperly segmented, or erroneous cell instances are reported despite no cells being present in the image. To assess SC-Track in generating reliable single cell linages from noisy CNN based cell segmentations, we repeated the tests with the uncorrected image segmentations which exhibited low levels of instance segmentation errors (Supplementary Table 3). Our results show a decrease in tracking accuracy for all the trackers tested (Fig. 5). Despite this, SC-Track gave the best overall performance, maintaining an average MOTA and IDF1 scores of > 0.9 and a CDF1 score of > 0.8. To further examine SC-Track's ability to overcome missing instances of cell segmentations, we generated a synthetic test dataset where cell instances were randomly removed at varying degrees (Supplementary Table 4). Our results show that SC-Track's cache matching algorithm can compensate for the loss of instance detections well and maintain an average IDF1 and MOTA score of > 0.9 in a dataset where 20% of all cell instances were missing (Fig. 5). Furthermore, despite increasing levels of missing instance detections, SC-Track can maintain its high reliability in detecting cell division events (Fig. 5).

To demonstrate that SC-Track can perform well in a diverse set of cell types and imaging conditions, we expanded our tracking benchmarks to a collection of publicly available microscopy datasets (Supplementary Table 5). We used the silver reference segmentation results from the Cell Tracking Challenge (CTC) because the CTC dataset contains a wide collection of timelapse microscopy images taken with a variety of imaging settings on various cancer cells of diverse morphologies[29]. The segmentation results from the CTC dataset are equally diverse ranging from nuclear masks to whole cell segmentations. We used the silver reference segmentation dataset since the segmentation results were derived from the best performing CNN models in the CTC[29]. Furthermore, the silver reference segmentations were accompanied by ground truth tracking results, making these datasets an impartial

real-life test to measure the generalisability of SC-Track's cell tracking algorithm. Our results show that SC-Track consistently displayed the best cell tracking performance as measured by MOTA and IDF1 scores for nearly all the CTC datasets (Fig. 5). Furthermore, utilising only the silver reference segmentation results, SC-Track can reliably detect cell division events in most of the CTC datasets (Fig. 5). These results provide evidence that SC-Track is an excellent general cell tracking algorithm that performs equally well on a variety of cell segmentation types and can maintain its high cell tracking performance under challenging conditions, including situations where the cell segmentation dataset exhibits high levels of detection loss.

Instance classification smoothing of single cell tracks and runtime evaluations.

When performing multi-class instance segmentations, it is often observed that the classifications of objects occasionally switch especially when the features exhibited by objects being detected does not completely fit into a particular class or suboptimal imaging conditions lead to misclassifications of detected objects. The inherent noise in the cell classifications can pose a problem if accurate classifications of cellular states are important, such as in the quantification of cell cycle phases in an image time series[28]. To overcome this inherent problem, we developed a TrackTree Class Smoothing (TCS) algorithm that employs a probabilistic cached class smoothing approach to help accurately identify cell phase transition points. To evaluate the utility of SC-Track's TCS algorithm, we measured the F1 scores of our custom trained StarDist model used to classify our test dataset on the various cell cycle phases predicted from the fluorescent PCNA signal (Fig. 6). The results indicate that TCS can improve the average F1 classification scores across all cell classes.

Finally, we conducted runtime tests for SC-Track to determine how long SC-Track takes to generate single cell tracks from cell segmentations. We measured the time taken to analyse cell segmentations from microscopy timelapse series of varying lengths (50-500 frames) and compared it with TrackMate, Deepcell-tracking, and pcnaDeep. Our results show that when working with small imaging datasets, SC-Track had the best performance (Fig. 6). However, the processing speed significantly decreased with increasing number of frames (Fig. 6). This was primarily caused by the

increase in duration required to load the timelapse microscopy images prior to the generation of single cell linages by SC-Track.

Discussion

In this study, we introduced SC-Track, a novel cell tracking algorithm that employs a hierarchical cell tracking methodology based on biological observations of cell division and cell movement dynamics. We show that SC-Track can generate highly accurate single cell tracks from both nuclear and cell segmentations of diverse morphologies and imaging conditions. To better assess the ability of cell trackers to accurately detect cell division events, we introduced a new metric called the Cell Division F1 (CDF1) score. Using this measure, SC-Track showed the best performance in detecting cell division events under all conditions tested. This was achieved without the finetuning of tracking parameters making SC-Track a desirable general cell tracking solution. Furthermore, its hierarchical probabilistic cache-cascade model can tolerate false or missing cell segmentations caused by the stochastic nature of CNNs, reducing the need for extensive time consuming manual corrections of image segmentations. In addition, we implemented a cache smoothing algorithm to help reduce the stochastic noise in cell classifications from CNNs while increasing the accuracy of the cell classifications of segmented cells in a time series. All these functionalities were achieved in a computationally efficient manner, allowing SC-Track to be run reliably without requiring access to a high-performance computing cluster.

In summary, SC-Track provides a solution to a longstanding problem involving the use of CNNs in the automated segmentation and classification of cells from timelapse microscopy images. To facilitate easy integration of SC-Track into image analysis pipelines that require its functionalities, SC-Track can generate accurate single cell tracks by using features extracted from cell segmentation masks only.

Materials and methods

Calculating similarity index when connecting segmented cells between frames

When there is more than one segmented cell overlapping with the previous frame, SC-Track will select the segmented cell with the highest similarity value with the segmented cell in the previous frame. SC-

Track will employ the following formula to determine the similarity value of the possible candidate pairs between frames:

$$S_i = \{P_1, P_2, \dots P_n\}, \text{where } P_n = (x_n, y_n)$$

$$sm_{i,j} = IoU(S_i, S_j) + Dis(S_i, S_j) + Sps(S_i, S_j) + Sas(S_i, S_j) + \Delta(S_i, S_j)$$

S_i represents the set of contour points in a 2D space defied by x_n, y_n for points $P_{1\to n}$ of a cell. $sm_{\mathrm{i,j}}$ represents the similarity index between the segmented cell i in the previous frame and the segmented cell j in the subsequent frame. Dis is the calculated distance between the centroid of the segmented cell i in the previous frame and the centroid of the segmented cell j in the current frame. IoU represents the intersection over union of the contours of cells i and j. Sps represents the shape similarity value [30], and Sas represents the area similarity of the two cells. $\Delta(S_i, S_j)$ represents additional supplementary features, such as the similarity in the variance or total intensity of fluorescent signals from segmented cells. To calculate $IoU(S_i, S_j)$, $Dis(S_i, S_j)$, $Sps(S_i, S_j)$, $Sas(S_i, S_j)$, and $\Delta(S_i, S_j)$, the following formula was employed:

$$IoU(S_i, S_j) = \frac{intersection(S_i, S_j)}{union(S_i, S_j)}$$

$$Dis(S_i, S_j) = \frac{1}{10^{-5} + \sqrt{(s_{ix} - s_{jx})^2 + (s_{iy} - s_{jy})^2}}$$

$Sps(S_i, S_j) = \sum_{\mathrm{n=1\dots7}} \left| \frac{1}{m_n^{S_i}} - \frac{1}{m_n^{S_j}} \right|$, where $m_n^{S_i}$ represents the seven Hu Moments.

$$Sas(S_i, S_j) = \frac{min(S_{i_{Area}}, S_{j_{Area}})}{max(S_{i_{Area}}, S_{j_{Area}})}, \text{where } S_{n\ Area} = \frac{1}{2}\left|\sum_{i=1}^{n}(x_i \cdot y_{i+1} - y_i \cdot x_{i+1})\right|$$

$\Delta(S_i, S_j) = \{\frac{mean(s_i)}{mean(s_j)}, \frac{var(s_i)}{var(s_j)}\}$, where $mean(s_i) < mean(s_j)$ and $var(s_i) < var(s_j)$

Bounding box expansion method for increasing candidate search area used to identify linked cells in adjacent frames for cell tracking

If there are no cells overlapping in the segmented area of a cell from the previous frame, SC-Track will expand its search area to search for potential candidates. The expansion of the search area utilises the bounding box of the segmented cell which is expended with the following formula:

$$Bc = Pos(x_1, x_2, y_1, y_2), where\ x_1 < x_2\ and\ y_1 < y_2$$

$$Ec = \alpha \cdot Bc$$

$$= Pos(x_1 - \alpha \cdot (x_2 - x_1), x_2 + \alpha \cdot (x_2 - x_1), y_1 - \alpha \cdot (y_2 - y_1), y_2 + \alpha \cdot (y_2 - y_1))$$

Bc represents the bounding box a cell. *Pos* represents the position of the bounding box with the minimum value of the segmented cell in the x axis and y axis represented by x_1 and y_1 while the maximum value as x_2 and y_2 respectively. Ec represents the expanded bounding box where potential cell candidates located in the current frame can be matched to the previous frame, α represents the coefficient for the expansion of the bounding box. By default, α is set to 1.

Benchmarking criteria and performance evaluation of cell tracking and classification accuracy

To evaluate the performance of SC-Track in accurately tracking segmented cells, we used performance measures established in the Multiple Object Tracking (MOT) framework which includes *IDF1* [24] and *MOTA*[22,23]. *IDF1* measures how long a tracker accurately identifies the correct segmented cells over a time series. It represents the ratio of correctly identified detections over the average number of ground-truth and computed detections[24]. *IDF1* is computed from the following formula:

$$IDF1 = \frac{2IDT}{2IDTP+IDFP+I}, \text{ where } IDP = \frac{IDTP}{IDTP+IDFP} \text{ and } IDR = \frac{IDTP}{IDTP+IDFN}$$

IDP represents the identification precision of the tracker which is computed as the average ratio of accurately identified true positives divided by the sum of accurately identified true positives and inaccurately classified false positives. *IDR* represents the identification recall which is computed as the average ratio of accurately identified true positives divided by the sum of accurately identified true positives and failed detections of each single cell track.

The multiple objects tracking accuracy (*MOTA*) measures the overall accuracy of the tracker performance using by measuring how often a mismatch occurs between the tracking results and the ground-truth[22,23]. This is obtained by computing the total number errors for false positives (*FP*), missed

targets (*FN*) and identity switches (*IDsw*) normalised over the total number of ground-truth (*GT*) tracks. This measure is computed using the following formula:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDsw_t)}{\sum_t GT_t}$$

To evaluate the reliability of the class smoothing algorithm, we employed the cell classification F1 score. The F1 score is calculated with the following formula:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Re}, \text{ where } Precision = \frac{TP}{TP+FP} \text{ and } Recall = \frac{TP}{TP+FN}$$

To measure SC-Track's ability to track cell division events, we have introduced a new indicator Cell Division F1 score (CDF1), which is calculated as:

$$CDF1 = \frac{2CDTP}{2CDTP + CDFP + CDFN}$$

$CDTP$ indicates a true positive cell division event, where both daughter cells of a cell division event are accurately identified and assigned to the correct TrackTree. $CDFP$ indicates false positive cell division event, where daughter cells are incorrectly assigned to a TrackTree and classified as a cell division event. $CDFN$ indicates a false negative cell division event, where a cell division event occurred but is not detected or the mother daughter cells were inaccurately assigned to the wrong TrackTree. The cell tracking outputs used to benchmark the tracking results can be obtained from Zenodo: https://zenodo.org/record/8284987. The python scripts used to analyse the cell tracking results can be found in GitHub: https://github.com/chan-labsite/SC-Track-evaluation.

Generation of in-house development and testing datasets

Two cell lines with distinct morphological appearances were used to generate the imaging data used in the development and testing of SC-Track. hTERT-RPE1 cells endogenously tagged with fluorescent mScarlet-PCNA were grown in DMEM/F-12 (Sigma, D6421) supplemented with 10% FBS (ExCell Bio, FSP500), 1× GlutaMAX (Gibco, 35050-061), 7.5% sodium bicarbonate (Sigma). MCF10A cells endogenously tagged with fluorescent mScarlet-PCNA were grown in DMEM/F-12 (Sigma,D6421) supplemented with 5% heat inactivated horse serum (Biological Industries, 04-124-1A), 1× GlutaMAX (Gibco, 35050-061), 10 μg/ml insulin (Biological Industries, 41-975-100), 10 ng/ml cholera toxin

(Sigma-Aldrich, #C-8052), 20 ng/ml EGF-β (Thermo Fisher, PHG0311), 0.5 mg/ml Hydrocortisone (MCE, HY-N0583). These cells were seeded in 8-Well chambered glass bottom slides (Cellvis, C8-1.5H-N) for two days before being imaged under a Nikon Ti2 inverted widefield fluorescence microscope equipped with a Lumencor Sola SE 365 as a light source. The cells were placed in an Okolab stage incubator (OKO) at 37°C with 5% CO2, and 80% humidity. The cells were observed under a 20× plan apo objective (NA 0.75) and images were captured using a Photometrics Prime BSI camera with a pixel resolution of 2048×2048. The following filter sets were used (mCherry: 560/40 nm EX, 585 nm BS, 630/75 nm EM). A single widefield image was taken in the mCherry channel (1% power, 200ms) at each stage at 5-minute intervals for up to 48h. A DIC image was captured at each time point (5% power, 100ms).

The timelapse microscopy images used as the development dataset for SC-Track was generated from as cells cultured under the conditions described above. The images were saved as individual multi-frame TIFF files. Four timelapse movies with varying cell densities per frame was generated within our lab (Supplementary Table 1). These datasets were automatically segmented using a custom pre-trained model of StarDist[12] and manually corrected using the VGG Image Annotator (VIA)[31] to remove false and inaccurate classifications. The annotated files contained two sets of information: the cell contour information and the "cell cycle phase" class information. The contour information was converted into a mask with values ranging from 1 to 255. The uncorrected and corrected mask images, along with the original mCherry channel image, constitute the datasets used to finetune the tracking parameters of SC-Track.

The timelapse microscopy images used in the testing dataset were generated under the conditions described above. In total, three RPE1 microscopy timelapse images and a two MCF10A microscopy timelapse images that were automatically segmented using our custom trained StarDist model and manually corrected to ensure accuracy of the instance segmentations, cell classifications and identity of single cell linages (Supplementary Table 2). The imaging conditions used were as described above with a sampling frequency of 5 minutes. To test the reliability of SC-Track to accurately track segmented cells with missing or false positive instances, we utilised the uncorrected segmentations of the testing dataset (Supplementary Table 3). In addition, to assess how SC-Track can cope with varying

levels of missing cell segmentations, we randomly deleted additional segmented cells from each frame to varying degrees to simulate higher levels of missed segmentations (Supplementary Table 4). The in-house generated segmentation masks, custom trained StarDist model, and ground truth tracking results used in testing SC-Track can be obtained from Zenodo: https://zenodo.org/record/8284987.

Cell Tracking Challenge dataset

To test the universal functionality of SC-Track to accurately generate single cell linages from a variety of cell types and segmentation modes, we used the silver reference segmentation results from the Cell Tracking Challenge[29]. The silver reference datasets represent the uncorrected segmentation results obtained from CNNs applied on a diverse variety of mammalian cell lines of different morphological appearances and imaging conditions (Supplementary Table 5). SC-Track with default settings was used to analyse the silver reference segmentation masks to generate single cell track linages. The single cell tracks generated by SC-Track was then compared with the accompanying ground truth tracking data provided by Cell Tracking Challenge to benchmark the reliability of SC-Track. The silver reference masks and ground truth tracking results were obtained from the Cell Tracking Challenge website (http://celltrackingchallenge.net/2d-datasets/).

Generation of single cell linages from segmentation masks

The segmentation results from the various evaluation datasets were used to measure the cell tracking performance of SC-Track and three other trackers pcnaDeep[28], Deepcell-tracking[27], and TrackMate[25,26]. For cell tracking experiments involving in-house generated testing datasets, the segmentation results in the form of a VGG image annotator (VIA2) compatible JSON file containing cell cycle class information of each segmented cell was used[31]. The data in the JSON files were read directly by SC-Track and pcnaDeep to generate the cell linage tables. The cell segmentation data in the JSON files were converted into greyscale multi-TIFF image files prior to being read by TrackMate and Deepcell-tracking as both software packages lack the function to directly read JSON files. To generate single cell linages from the Cell Tracking Challenge dataset, the silver reference segmentation results in the form of a greyscale TIFF image series were used for SC-Track, TrackMate and Deepcell-tracking.

We could not perform the cell tracking experiments with the Cell Tracking Challenge dataset on pcnaDeep as it requires cell cycle class information to function[28].

Default tracking settings were applied to SC-Track, pcnaDeep and Deepcell-tracking. For TrackMate, the Lap tracker algorithm was used with default tracking settings. The scripts used to generate the cell tracking results can be obtained from GitHub: https://github.com/chan-labsite/SC-Track-evaluation.

Automated cell cycle class correction testing

To evaluate the cell class correction function of SC-Track, the same testing dataset with the uncorrected cell classifications obtained from our custom pre-trained StarDist model was utilised (Supplementary Table 3). Ground truth cell cycle classifications were obtained by manual correction of the automated annotations were used to compute the F1 scores for individual cell cycle classifications in the timelapse image series. The JSON file containing the raw uncorrected cell segmentations and the cell cycle classification data used to compute the F1 results can be obtained from Zenodo: https://zenodo.org/record/8284987. The scripts used to compute the F1 scores of individual cell cycle phases can be obtained from GitHub: https://github.com/chan-labsite/SC-Track-evaluation.

Runtime and multi-platform compatibility testing

We conducted compatibility tests on Windows, Linux, and macOS platforms. In addition, we performed runtime efficiency tests specifically on the Windows platform. All tests were performed using the same dataset and repeated three times. The Windows platform was configured with an AMD R7 3700X CPU, RTX 2080 GPU, and 16GB of RAM. The Linux platform was configured with an Intel i7 11800H CPU, RTX 3050Ti GPU, and 16GB of RAM. The macOS platform was configured on a 2021 MacBook Pro equipped with a M1 processor, and 8GB of RAM.

Funding

This work was supported by funds from Zhejiang University and the Fundamental Research Funds for the Central Universities (2019QN30001) to KYC. SSX postdoctoral fellowship is funded by ZJE Institute of Zhejiang University.

Author contributions

KYC conceived the study. KYC, CL, SSX and JW conceived the experiments. CL, SSX and JW conducted the experiments. KYC, CL and SS wrote the manuscript. KYC, CL, SSX, JW and SS analysed the results. All authors reviewed the manuscript.

Data availability

The raw and corrected cell segmentation results, raw cell tracking outputs, and the raw analysis results can be obtained from Zenodo: https://zenodo.org/record/8284987. The silver reference masks and ground truth tracking results were obtained from the Cell Tracking Challenge website (http://celltrackingchallenge.net/2d-datasets/).

Code availability

A python implementation of SC-Track with its corresponding usage documentations is available at GitHub: https://github.com/chan-labsite/SC-Track. The scripts used to analyse the raw data and to generate the figures presented in this manuscript are available at GitHub: https://github.com/chan-labsite/SC-Track-evaluation.

References

1. Skylaki, S., Hilsenbeck, O. & Schroeder, T. Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol.* **34**, 1137–1144 (2016).
2. Cooper, S. & Bakal, C. Accelerating Live Single-Cell Signalling Studies. *Trends Biotechnol.* **35**, 422–433 (2017).
3. Min, M., Rong, Y., Tian, C. & Spencer, S. L. Temporal integration of mitogen history in mother cells controls proliferation of daughter cells. *Science (80-.).* **368**, 1261–1265 (2020).

4. Spencer, S. L. *et al.* The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit. *Cell* **155**, 369–383 (2013).
5. Cappell, S. D. *et al.* EMI1 switches from being a substrate to an inhibitor of APC/CCDH1 to start the cell cycle. *Nature* **558**, 313–317 (2018).
6. Arora, M. *et al.* Rapid adaptation to CDK2 inhibition exposes intrinsic cell-cycle plasticity. *Cell* **186**, 2628-2643.e21 (2023).
7. Barr, A. R. *et al.* DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nat. Commun.* **8**, 14728 (2017).
8. Meijering, E. A bird's-eye view of deep learning in bioimage analysis. *Computational and Structural Biotechnology Journal* vol. 18 2312–2325 at https://doi.org/10.1016/j.csbj.2020.08.003 (2020).
9. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
10. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **9351**, 234–241 (2015).
11. Moen, E. *et al.* Deep learning for cellular image analysis. *Nature Methods* vol. 16 1233–1246 at https://doi.org/10.1038/s41592-019-0403-1 (2019).
12. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **11071 LNCS**, 265–273 (2018).
13. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods 2022 1912* **19**, 1634–1641 (2022).
14. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
15. Bilodeau, A. *et al.* Microscopy analysis neural network to solve detection, enumeration and segmentation from image-level annotations. *Nat. Mach. Intell. 2022 45* **4**, 455–466 (2022).
16. Meijering, E. Cell segmentation: 50 Years down the road [life Sciences]. *IEEE Signal Process.*

Mag. **29**, 140–145 (2012).

17. Caicedo, J. C. *et al.* Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images. *Cytom. Part A* **95**, 952–965 (2019).

18. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods 2020 181* **18**, 100–106 (2020).

19. Xie, S. & Skotheim, J. M. A G1 Sizer Coordinates Growth and Division in the Mouse Epidermis. *Curr. Biol.* **30**, 916-924.e2 (2020).

20. Fantes, P. A., Grant, W. D., Pritchard, R. H., Sudbery, P. E. & Wheals, A. E. The regulation of cell size and the control of mitosis. *J. Theor. Biol.* **50**, 213–244 (1975).

21. Guerrero-Pena, F. A. *et al.* Multiclass Weighted Loss for Instance Segmentation of Cluttered Cells. *Proc. - Int. Conf. Image Process. ICIP* 2451–2455 (2018) doi:10.1109/ICIP.2018.8451187.

22. Bernardin, K. & Stiefelhagen, R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP J. Image Video Process. 2008 20081* **2008**, 1–10 (2008).

23. Leal-Taixé, L. *et al.* MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. (2015).

24. Ristani, E., Solera, F., Zou, R., Cucchiara, R. & Tomasi, C. Performance measures and a data set for multi-target, multi-camera tracking. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **9914 LNCS**, 17–35 (2016).

25. Ershov, D. *et al.* TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat. Methods 2022 197* **19**, 829–832 (2022).

26. Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).

27. Moen, E. *et al.* Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning. *bioRxiv* 803205 at https://doi.org/10.1101/803205 (2019).

28. Gui, Y. *et al.* pcnaDeep: a fast and robust single-cell tracking method using deep-learning mediated cell cycle profiling. *Bioinformatics* **38**, 4846–4847 (2022).

29. Maška, M. *et al.* The Cell Tracking Challenge: 10 years of objective benchmarking. *Nat.*

Methods 2023 207 **20**, 1010–1020 (2023).

30. Hu, M. K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. Inf. Theory* **8**, 179–187 (1962).

31. Dutta, A. & Zisserman, A. The VIA Annotation Software for Images, Audio and Video. in *Proceedings of the 27th ACM International Conference on Multimedia* 2276–2279 (ACM, 2019). doi:10.1145/3343031.3350535.

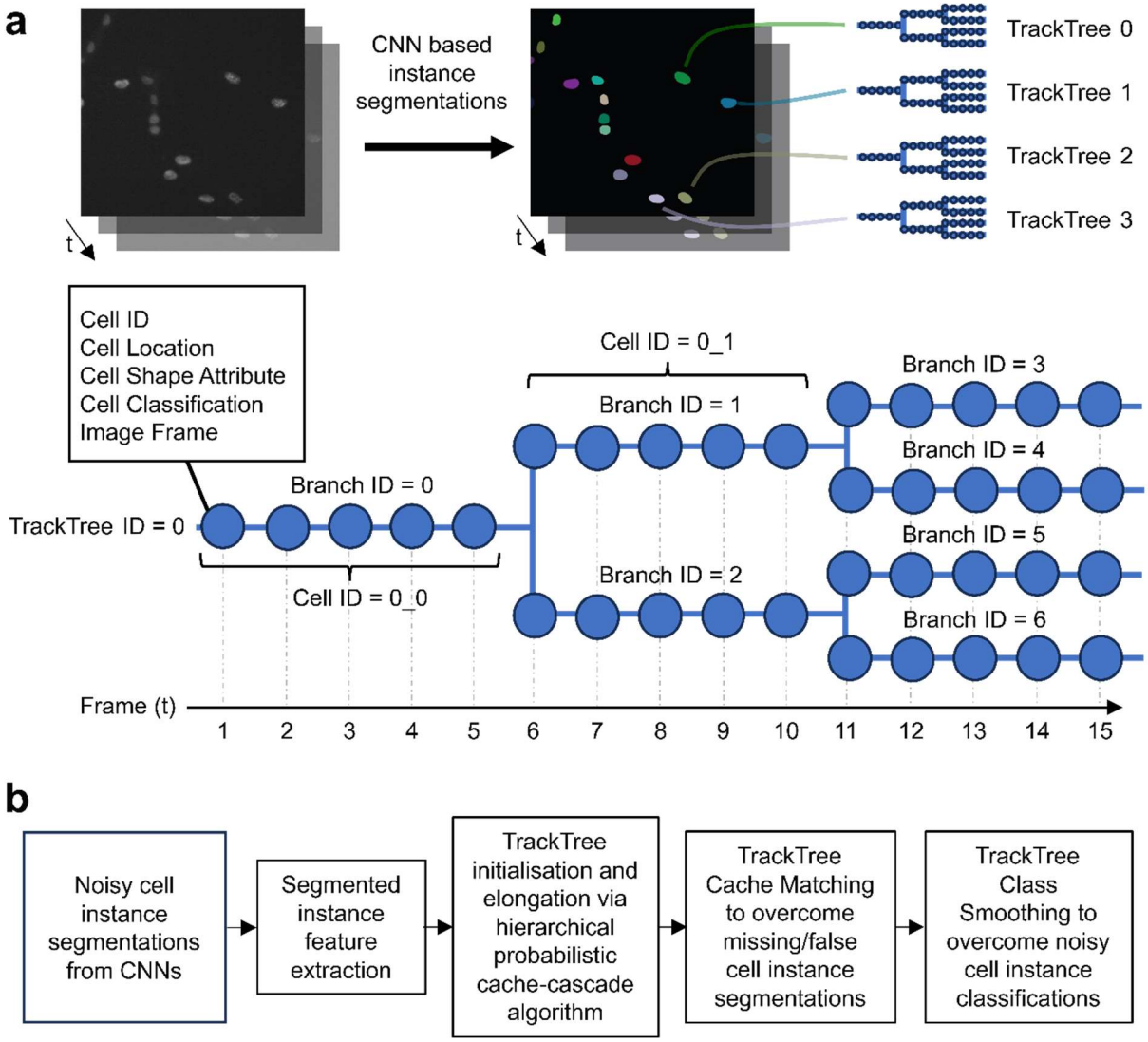


Fig. 1: Schematic illustration providing an overview of SC-Track, the TrackTree data structure and analysis pipeline.

a, A summary of the TrackTree data structure. Each linked segmented cell is tracked in a TrackTree. A node in a TrackTree branch represents an instance of the segmented cell in a particular frame with its accompanying cell segmentation information. A branching of a TrackTree represents a cell division event. **b**, A simplified overview of the analysis pipeline of SC-Track. Instance segmentation of cells from each frame is sequentially added to their respective TrackTrees. The assignment of each instance segmentation is determined by the hierarchical probabilistic cache-cascade model of SC-Track. If there are cell classification information contained in the TrackTrees, SC-Track will employ the TrackTree Class Smoothing (TCS) algorithm to correct the noisy cell classifications.

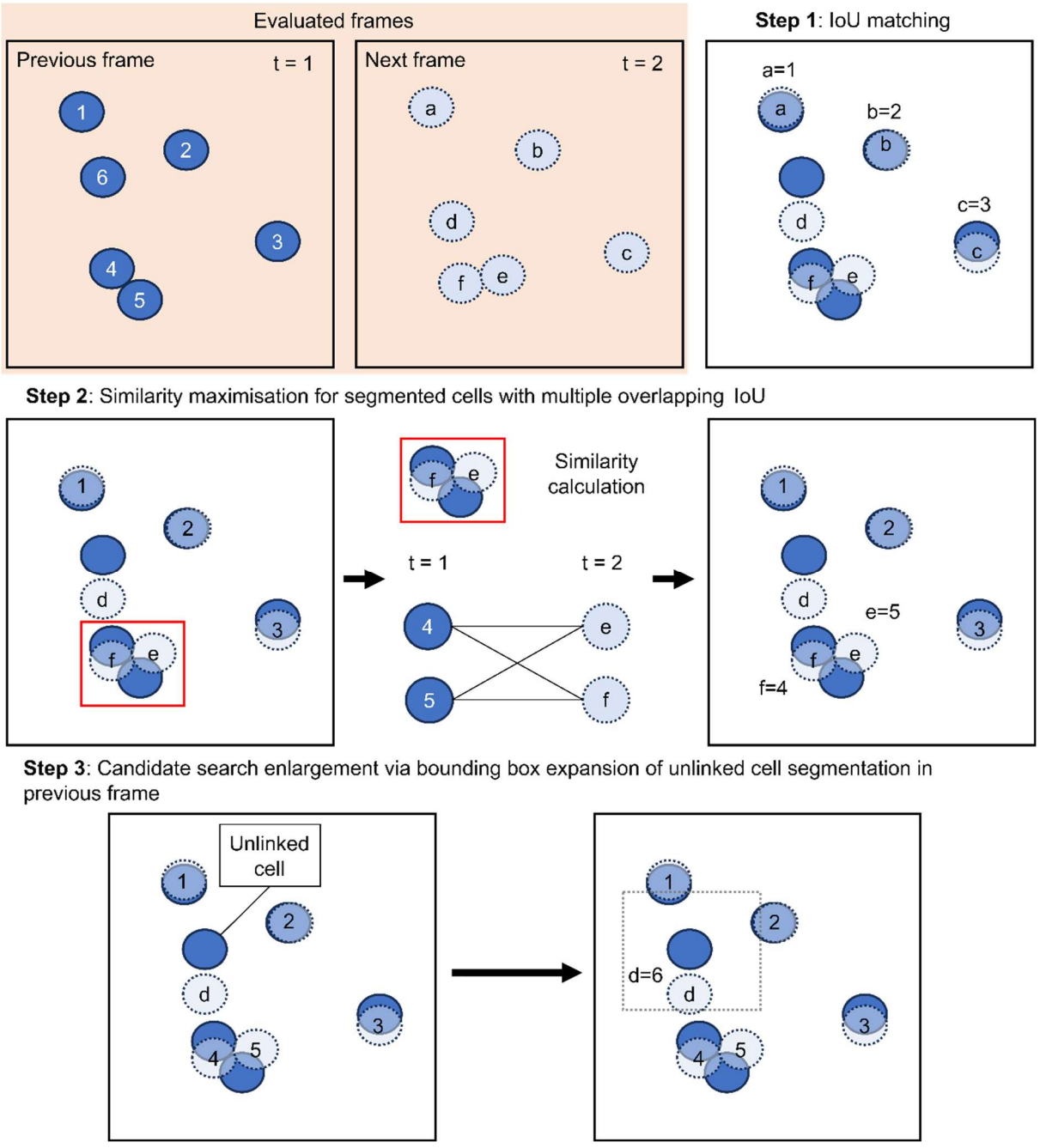


Fig. 2: Schematic illustration summarising the hierarchical tracking approach for single cell tracking.

SC-Track employs a hierarchical cell tracking approach to minimise computational costs. The initial linking of segmented cells between frames is initially determined by the overlap between the segmented cells of the preceding and subsequent frame. If there is only one cell segmentation overlapping, the segmented cell in the subsequent frame is automatically linked to the respective TrackTree of the overlapped cells in the preceding frame. When there are multiple overlapping cells, the identification

of the linked cells will be determined by the similarity value of the overlapping cells of the subsequent frame with the segmented cell in the preceding frame. If no overlapping candidate segmented cell was identified with a preceding segmented cell, the bounding box of the preceding cell was identified, the bounding box of the segmented cell will be expended to identify possible candidates.

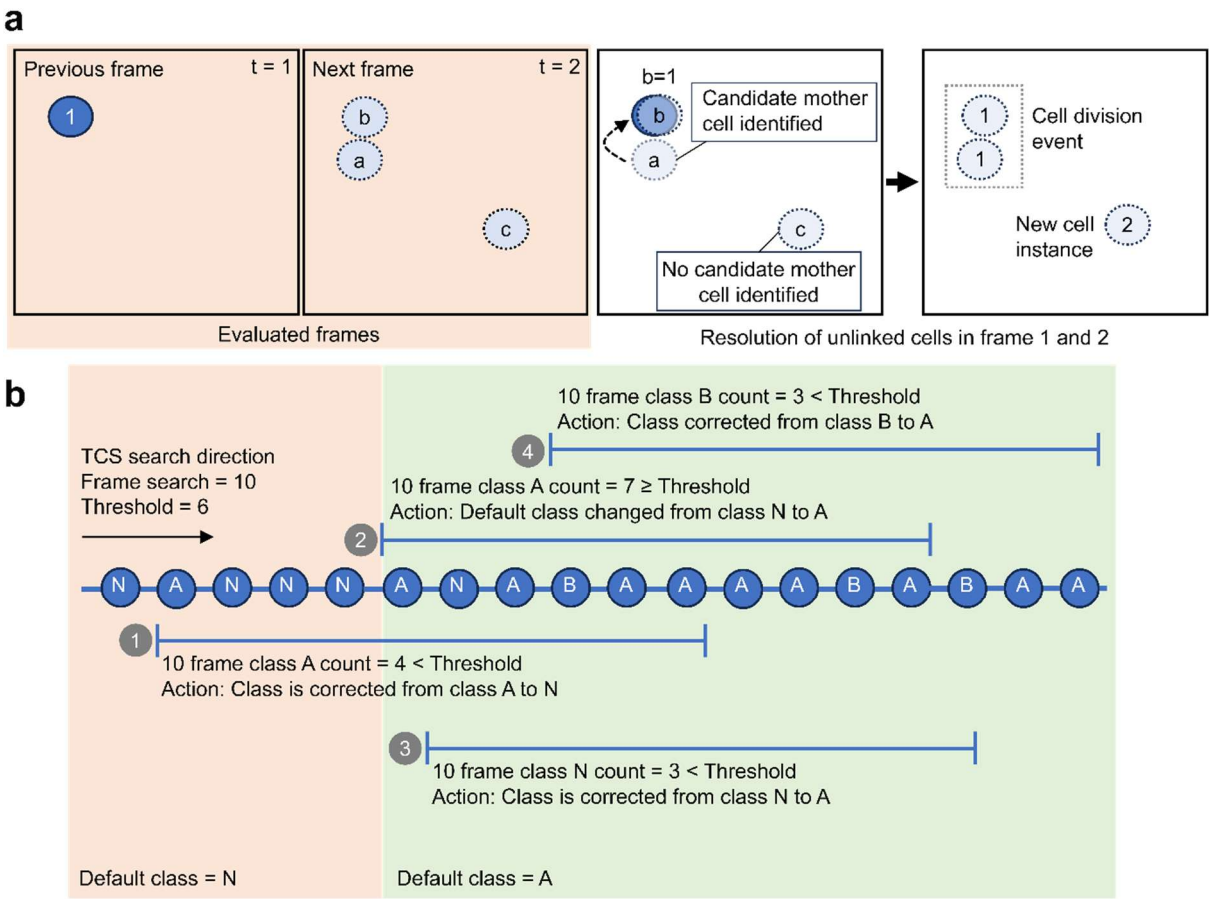


Fig. 3: Schematic illustration describing SC-Track algorithm in identifying cell division events and TrackTree Class Smoothing (TCS).

a, When a new segmented cell instance that cannot be linked to available TrackTrees is identified, SC-Track will attempt to determine if a cell division event has occurred. If a compatible candidate mother cell is identified in the preceding frame, the new segmented cell instance will be added to the corresponding TrackTree and a cell division event is recorded. If no compatible mother cell is identified, SC-Track will assume that this is a new segmented cell instance is due to a recent appearance of a cell into the microscope field of view and a new TrackTree is initialised. **b**, When a multi-class cell segmentation is performed, it is often observed that erroneous cell classifications would occur stochastically. The TCS algorithm employs a probabilistic cached search algorithm to determine if a class switch has occurred for the respective cell in a time series.

a

Dataset	SC-Track			pcnaDeep			TrackMate			Deepcell-tracking		
	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1
RPE1-01	**0.9897**	**1.0000**	**0.9647**	0.9498	0.9574	0.6378	0.7543	0.9538	0.3077	0.9156	0.9833	-
RPE1-02	**0.9993**	**0.9999**	**0.9587**	0.9565	0.9982	0.5333	0.6602	0.6651	0.2286	0.9351	0.9898	-
MCF10A-01	**0.9936**	**0.9957**	**0.9063**	0.9727	0.9921	0.4957	0.8576	0.9471	0.0223	0.9616	0.9808	-
MCF10A-02	0.9711	**0.9977**	**0.9189**	**0.9822**	0.9782	0.6567	0.8780	0.9570	0.0000	0.9516	0.9665	-
RPE1-03	**0.9832**	**0.9952**	**0.9000**	0.9748	0.9946	0.3077	0.4835	0.6111	0.0235	0.9314	0.9764	-
Average	**0.9874**	**0.9977**	**0.9297**	0.9672	0.9841	0.5263	0.7267	0.8268	0.1164	0.9391	0.9794	-

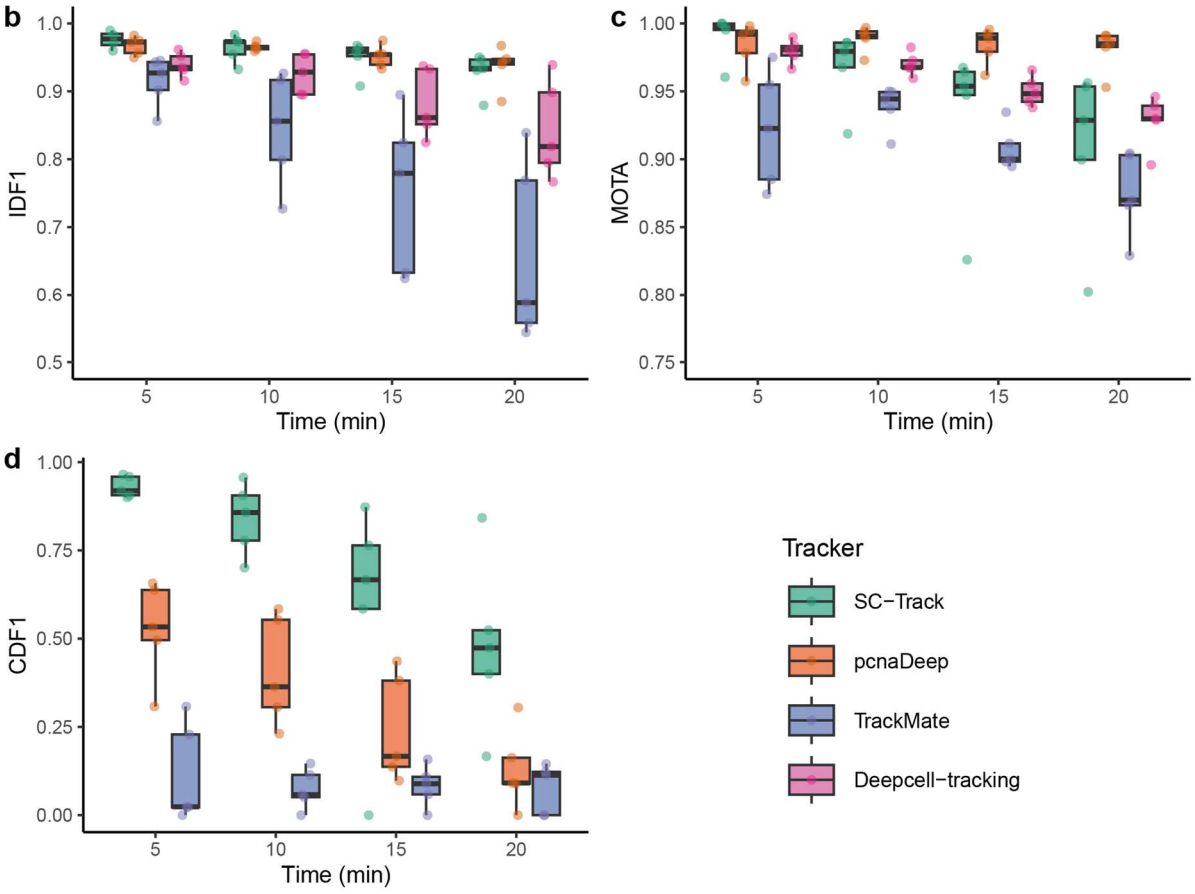
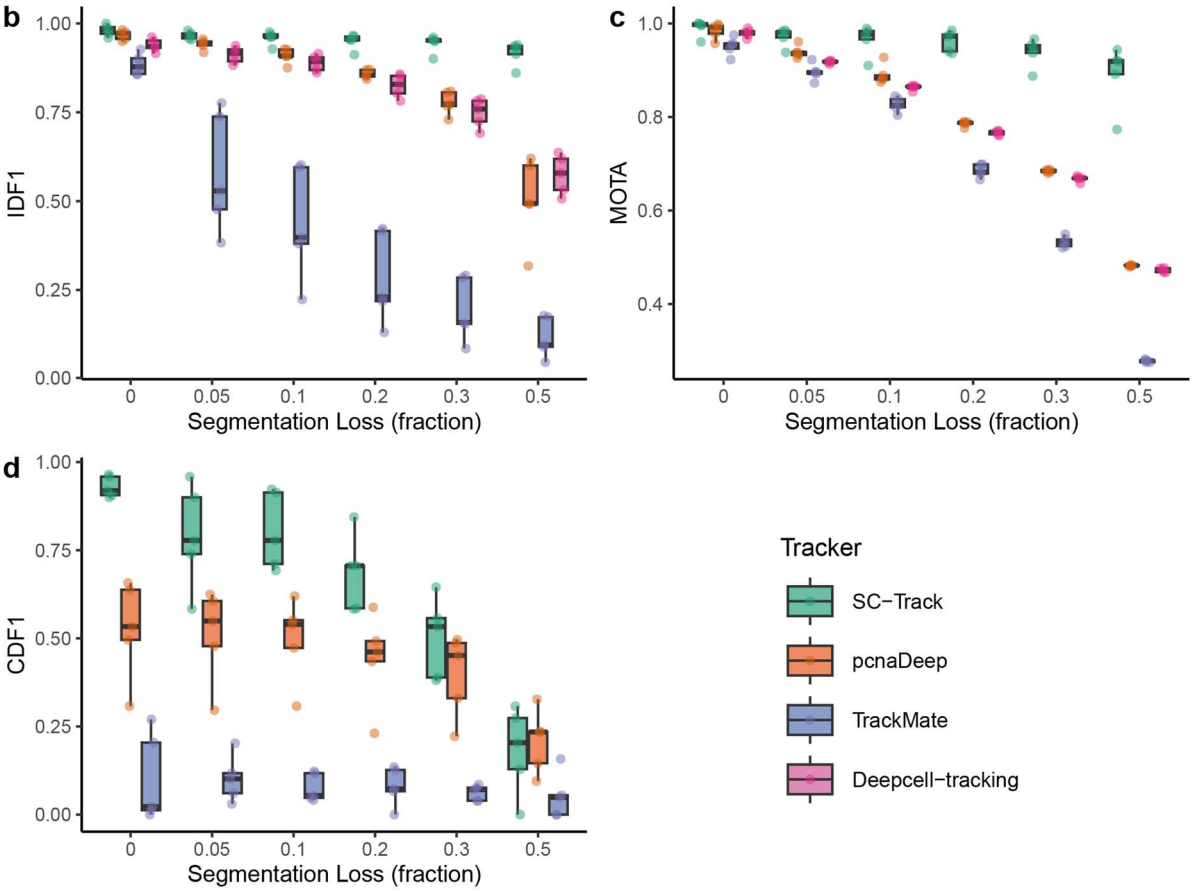


Fig. 4: Evaluation metrics of the cell tracking accuracy based on ground truth segmentations.

a, Table showing the IDF1, MOTA and CDF1 scores of all four trackers. The best scores for each respective dataset and the best average score are highlighted in bold. **b-d**, Box plots of IDF1, MOTA and CDF1 scores for all four cell trackers in varying imaging time intervals. Each point displayed on the boxplots represent the respective scores of the five test datasets. The line in the boxplot represents the median. The results for Deepcell-tracking CDF1 scores were not included in (**d**) as the tracker failed to detect any cell division instances in all the datasets tested.

a

Dataset	SC-Track			pcnaDeep			TrackMate			Deepcell-tracking		
	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1
RPE1-01	0.9483	0.9596	**0.9072**	**0.9533**	**0.9816**	0.3913	0.7542	0.8760	0.1929	0.9290	0.9736	-
RPE1-02	**0.9504**	0.9431	**0.7636**	0.9136	**0.9796**	0.4795	0.7461	0.9475	0.0881	0.9190	0.9708	-
MCF10A-01	0.9324	**0.9688**	**0.8406**	0.9420	0.9570	0.3455	0.7548	0.8291	0.1081	**0.9423**	0.9552	-
MCF10A-02	**0.9528**	**0.9615**	**0.9714**	0.9443	0.9534	0.2857	0.7870	0.8741	0.0909	0.9476	0.9501	-
RPE1-03	0.9533	0.9898	**0.9565**	**0.9670**	**0.9945**	0.3077	0.7708	0.9633	0.0808	0.8899	0.9580	-
Average	**0.9474**	0.9646	**0.8879**	0.9440	**0.9732**	0.3619	0.7626	0.8980	0.1122	0.9256	0.9615	-



e

Dataset	Temporal res (min)	Image	Mask	Division events	SC-Track			TrackMate			Deepcell-tracking		
					IDF1	MOTA	CDF1	IDF1	MOTA	CDF1	IDF1	MOTA	CDF1
DIC-C2DH-HeLa	10	DIC	Cell	2	**0.9922**	**0.9922**	**0.8000**	0.5682	0.6392	0.0000	0.2060	-0.0196	-
PhC-C2DH-U373	15	Phase	Cell	0	**1.0000**	**1.0000**	-	0.7187	0.9542	-	0.7611	0.6039	-
Fluo-C2DL-MSC	20	Fluor	Cell	1	**0.9627**	**0.9907**	**0.6667**	0.4447	0.6853	0.0000	0.5885	0.5991	-
Fluo-N2DH-GOWT1	5	Fluor	Nuclear	1	**0.9908**	**0.9990**	0.0000	0.9712	0.9476	0.0000	0.9024	0.8551	-
Fluo-N2DH-SIM+	29	Fluor	Nuclear	16	**0.9548**	**0.9856**	**0.9677**	0.6895	0.6337	0.1509	0.7474	0.9248	-
PhC-C2DL-PSC	10	Phase	Cell	44	**0.9886**	**0.9997**	**0.9890**	0.9032	0.9446	0.3007	0.8789	0.9911	-
Fluo-N2DL-HeLa	30	Fluor	Nuclear	56	0.8799	0.9717	**0.9910**	**0.8813**	0.8548	0.0606	0.8069	**0.9888**	-
BF-C2DL-MuSC	5	BF	Cell	5	**1.0000**	**1.0000**	**0.3333**	0.1429	-0.3480	0.0800	0.5402	0.0958	-
				Average	**0.9711**	**0.9924**	**0.6782**	0.6650	0.6639	0.0846	0.6789	0.6299	-

Fig. 5: Evaluation metrics of the cell tracking accuracy based on diverse cell segmentation qualities and modalities.

a, Table showing the IDF1, MOTA and CDF1 scores of tracking results based on raw uncorrected cell segmentations obtained from a custom trained StarDist model. The best scores for each respective

dataset and the best average score are highlighted in bold. **b-d**, Boxplots of IDF1, MOTA and CDF1 scores for all four cell trackers with varying levels of cell segmentation loss. Each point displayed on the boxplots represent the respective scores of the five test datasets. The line in the boxplot represents the median. The results for Deepcell-tracking CDF1 scores were not included in (**d**) as the tracker failed to detect any cell division instances in all the datasets tested. **e**, IDF1, MOTA and CDF1 test results for the Cell Tracking Challenge (CTC) silver reference dataset. We were unable evaluate pcnaDeep's cell tracking performance on the CTC dataset because pcnaDeep requires cell cycle data encoded in the cell segmentations to generate single cell tracks.

a

Dataset	Raw F1 values			TCS corrected F1 values		
	G1/G2	S	M	G1/G2	S	M
RPE1-01	0.9555	0.7749	0.6594	**0.9923**	**0.9780**	**0.7906**
RPE1-02	0.9492	0.8093	**0.7483**	**0.9793**	**0.9216**	0.6313
MCF10A-01	0.9305	0.6900	0.4688	**0.9779**	**0.9232**	**0.4803**
MCF10A-02	0.8814	0.6597	0.4793	**0.9538**	**0.7493**	**0.6895**
RPE1-03	0.9645	0.7950	0.2358	**0.9949**	**0.9626**	**0.6667**
Average	0.9362	0.7458	0.5183	**0.9797**	**0.9069**	**0.6517**

b

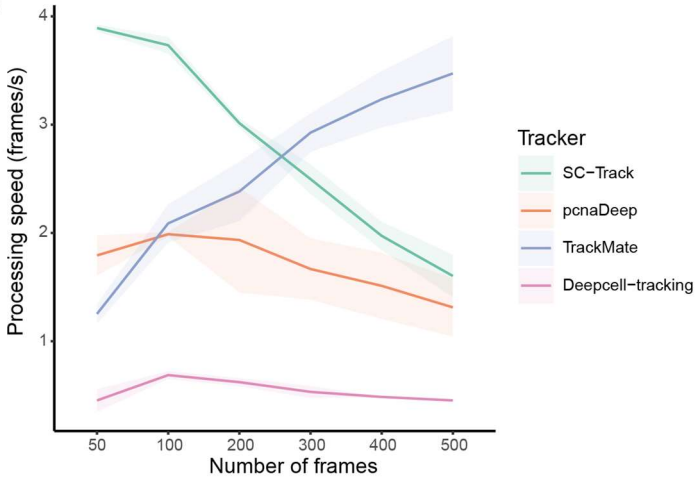


Fig. 6: Performance evaluation of SC-Track's TCS class correction algorithm and runtime evaluation comparisons.

a, F1 cell cycle classification test results obtained from raw StarDist cell classification predictions compared with TCS corrected cell classifications. The best scores for each respective cell classification and dataset are highlighted in bold. **b**, The average number of frames each tracker can process in one second is displayed in they y-axis while the x-axis represents the varying number of image frames were processed respectively. The solid line represents the average performance with the shaded area representing the 95% confidence interval for each cell tracker on three different computer systems running either Windows, Linux or macOS operating systems respectively.