Siamese Tracking of Cell Behaviour Patterns

Editors: Under Review for MIDL 2020

Abstract
Tracking and segmentation of biological cells in video sequences is a challenging problem, especially due to the similarity of the cells and high levels of inherent noise. Most machine learning-based approaches lack robustness and suffer from sensitivity to less prominent events such as mitosis, apoptosis and cell collisions. Due to the large variance in medical image characteristics, most approaches are dataset-specific and do not generalise well on other datasets.

In this paper, we propose a simple end-to-end cascade neural architecture able to model the movement behaviour of biological cells and predict collision and mitosis events. Our approach uses U-Net for an initial segmentation which is then improved through processing by a siamese tracker capable of matching each cell along the temporal axis. By facilitating the re-segmentation of collided and mitotic cells, our method demonstrates its capability to handle volatile trajectories and unpredictable cell locations while being invariant to cell morphology. We demonstrate that our tracking approach achieves state-of-the-art results on PhC-C2DL-PSC and Fluo-N2DH-SIM+ datasets and ranks second on the DIC-C2DH-HeLa dataset of the cell tracking challenge benchmarks.

Keywords: Mitosis, cell collision, segmentation, cell tracking, Siamese tracker

1. Introduction
Advancements in the field of machine learning (ML) have allowed for the automation of medical data analysis, with several advanced methods performing as good as humans on a range of tasks (e.g., Dalca et al. 2018; Chen et al. 2018). Example problems where deep learning (DL) methods have produce significant impact include detection and classification of tumours, (Coudray et al., 2018; Kamnitsas et al., 2017; Budinska et al., 2013; Kalaiselvi and Nasira, 2014), identification of new biomarkers in high-dimensional data (Budinska et al., 2013; Ravi and Hegadi, 2015; Rathi et al., 2011), among others.

The problem of medical data analysis becomes more challenging at the micrometer scale (e.g. pancreatic stem cells or cell nuclei), where it is difficult to visualise and process the data (Coudray et al., 2018; Saltz et al., 2018). Cell distinctive shapes and morphological traits are convoluted between low resolution artefacts, noise components from the microscope scanning device and lighting conditions (Swiderska-Chadaj et al., 2019). Figure 1 shows example images of cells for five different datasets from the IEEE International Symposium on Biomedical Imaging (ISBI) cell tracking challenge (Maška et al., 2014). As can be seen from these images, there are various challenges such as low signal-to-noise ratio, poor illumination, clutter of cells and occlusion, that make it difficult to accurately track and segment the cells with a naked eye.
In addition to the challenges outlined by Figure 1, biological cells also fail to conform to a predefined shape and there can be several shifting movement patterns that are hard to analyse or detect (Saltz et al., 2018). Examples of such behaviour are shown in Figure 4 where a parent cell gets split into two daughter cells (Mitosis) or two cells collide and appear as a single cell (Collision).

Some cell segmentation and tracking approaches include constructing temporal trajectories for cells (Yang et al., 2005), spatial correlation using Delaunay graphs (Nath et al., 2006), and watershed deconvolution with morphological operators (Sharif et al., 2012). Due to the large noise components of medical images, ML methods mediating dataset specific properties such as uneven illumination and lack of pixel-value normalisation tend to define many morphological dependent conditions for every cell type (Sharif et al., 2012; Lux and Matula, 2019). However, this process is sensitive to outliers and along with volatile positional changes, it makes tracking each individual cell over time more difficult. This is because a cell in the previous frame might change completely in the next.

The U-Net neural network (NN) introduced by Ronneberger et al. (2015) is among the most successful NN architectures in biomedical image segmentation for tasks such as tumour detection and living cell segmentation (Heller et al., 2019; Falk et al., 2019). Due to its U-shape and essential residual connections, U-Net has achieved state-of-the-art results in many challenges (Li et al., 2018; Dubost et al., 2017). However, despite its high performance, even U-Net struggles on images containing significant movements of cells and changes in their morphology, and in particular, fails to reliably detect cells that split or die (leave the field of view of the scanning device) (Christ et al., 2016). Lux and Matula (2019) combined U-Net with watershed deconvolution (Kachouie et al., 2008), and demonstrated improved performance on cell tracking datasets\textsuperscript{1}. However, this approach involves tuning of several parameters, i.e. several data-specific intermediate processing steps relating to cell size, erosion, staining, and other morphological traits so as to acquire reasonable cell shapes. Zhou et al. (2019) proposed using two U-Net architectures, one for segmenting cells and the other for detecting their centroids. Due to learn data specific network weights for cell detection, this approach is more resilient to outliers, however, it does not suggest a robust way to detect smaller cells that are falsely segmented as a single cell.

Lux and Matula (2019) used static area overlap for building the correspondence of cell trajectory in subsequent frames. Another approach uses level sets to follow the evolution of cells (Maška et al., 2014). However, due to the fluidic and erratic nature of cells, neither

\begin{figure}
\centering
\includegraphics[width=\textwidth]{example_images.png}
\caption{Example images of (a) Fluo-N2DL-HeLa (b) Fluo-N2DH-SIM (c) Fluo-N2DH-GOWT1 (d) PhC-C2DL-PSC (e) DIC-C2DH-HeLa datasets from the cell tracking challenge (Maška et al., 2014).}
\end{figure}

\textsuperscript{1} \url{http://celltrackingchallenge.net/}
Siamese Tracking of Cell Behaviour Patterns

These algorithms are able to model their real movement patterns. Siamese networks, first introduced in the work of Bromley et al. (1994) and adapted for Siamese Instance Search Tracking (SINT) by Tao et al. (2016), have shown to excel in generic object tracking (Bertinetto et al., 2016). Due to their robustness in matching under appearance variations, siamese methods have been used in segmentation of medical images (Spitzer et al., 2018). It is of interest to explore the applicability of siamese networks for tracking the cells.

In this paper, we propose a Tracking-Assisted Segmentation (TAS) approach by combining a U-Net architecture with a Siamese matching-based tracker. Our proposed method can easily model and detect the unstable biological cell movement activities such as mitosis and cell collisions that lead to false predictions. Further, through the use of watershed algorithm, the false predictions can be re-segmented to improve the overall cell tracking performance. The contributions of this paper can be summarised as follows.

- We augment segmentation by siamese tracking for improved temporal correspondence and re-segmentation of erroneous predictions.
- Our approach is more robust to morphology variations and explicitly models rare events such as mitosis, apoptosis and cell collisions.
- Our approach generalises well to different datasets outperforming state-of-the-art segmentation methods for biological cells on three benchmark datasets.

The source code will be made publicly available with the published version of this paper.

2. Methodology

The method proposed here is a two step process with input being the raw images and the final output being per pixel classification labels as shown in Figure 2. The pipeline of our approach can be understood from the pipeline shown in Algorithm 1. Below, we briefly discuss the various parts of the process.

Initial Segmentation. The initial segmentation of the cells is performed using three different approaches, namely TAS-general, TAS-intermediate and TAS-specialised. The three approaches differ in the sense that TAS-general does not involve any parameter tuning with respect to datasets, TAS-intermediate involves limited adjustments of parameters to account for the varying image properties, and TAS-specialised is properly tuned for every dataset.

TAS-general. As the name suggests, this is a generalised implementation of tracking-assisted segmentation approach that requires no tuning or adjustments with respect to any datasets. Recent methods use the watershed algorithm for splitting cell bodies into smaller cells (Sharif et al., 2012; Lux and Matula, 2019). However, this approach is very sensitive to noise and in order to keep the method general, we thus use the random walker algorithm, as explained in the work of Grady (2006). The output of the random walker, referred as the initial segmentation, is then fed into the tracker network to correct for mitosis, apoptosis and cell collisions.

For TAS-general, U-Net (Ronneberger et al., 2015) is explicitly trained on each dataset, 3.1, using data augmentation with additive noise, pixel value range shift and a cutoff on the

---

2. On https://gitlab.com/
Algorithm 1: Tracking-Assisted Segmentation (TAS)

**Input:** Sequence of frames  
**Output:** Per frame cell segmentation and tracking information

for frame in sequence do

Step 1: Initial segmentation
- Separate cell bodies from the background using U-Net
- Define cell identities using watershed or random walker

Step 2: Siamese tracking
- Track each cell in small intervals
- Identify biological cell behaviours:
  - Collision, if $match(cell_t, cell_{t-1}^{(1)}, cell_{t-1}^{(2)})$
  - Mitosis, if $match(cell_t, cell_{t+1}^{(1)}, cell_{t+1}^{(2)})$
  - Cell death, if $match(cell_t, None)$
- Re-segment cells using watershed

end

maximum value. The choice of this specific augmentation arises from the observation of the dataset images which are not consistent in a standard RGB value range and have low signal-to-noise ration as indicated by Maška et al. (2014). The network is trained with a learning rate of 0.001 using the Adam optimiser for 50 epochs. Each raw image is normalised based on its histogram to enhance contrast (Pizer et al., 1987) by a linear interpolation using the cumulative distribution of the histogram as explained in the work of Kim (1997). Further normalisation includes zero mean and unit variance transformation before feeding images to the U-Net for image segmentation. The U-Net predictions separate cell bodies from the background, and the Euclidean distance of the cell areas to the background is used as a threshold value, to limit the centroid areas of the cells. More details related to the experimental implementation are presented in Appendix B. Unlike the methods that use thresholds based on morphology (Lux and Matula, 2019), the pre-processing of the raw images and post-segmentation division of small cells are also generalised, and not data-specific.

TAS-intermediate. This approach relaxes the constraint of a general pre-processing method and uses the U-Net implementation of Ronneberger et al. (2015), pre-trained as in the work of Lux and Matula (2019). The same pre-processing steps are employed, and the model is manually fine-tuned to the image properties based on the type of cell image. To evaluate the effect of morphological boundary refinements on cell segmentation, the approach of defining unique cells from the initial segmentation, using the random walker algorithm, is kept the same as in TAS-general. This second variant of the initial segmentation process is referred as TAS-intermediate.

TAS-specialised. The third approach relaxes further the constraint of dataset-tuned segmentation and applies the water deconvolution method as described in the work of Lux and Matula (2019) instead of the random walker algorithm. Compared to TAS-intermediate, the
difference is that this implementation of our model also now uses several tuning parameters which are typically adjusted for every dataset. Thus, this implementation of TAS can be interpreted as a tuned implementation of TAS that is specialised for a certain dataset.

The output of the initial segmentation produces finely segmented cells where the actual cell behaviours might not be correctly expressed. A schematic representation related to this issue is shown in Figure 3 of Appendix C. To be able to correctly reason about cell movement over time, cells need to be correctly tracked.

**Siamese tracking.** The location of the cell in subsequent frames is identified using siamese tracking. For tracking purposes, a SiamFC tracker (Bertinetto et al., 2016), pre-trained on the GOT-10k dataset (Huang et al., 2019), is used. Tracking is done in the forward as well as the backward direction to predict the new location of a cell from the current frame to the previous and next frames. The cell segmentation is refined through tracking by detecting the occurrences of mitosis and collision events, using the same approach but in opposite directions along the temporal dimension.

The working of the tracking module is as follows. Let $I_t$ denote the $t^{th}$ frame in a sequence of length $T$, and $S_t = \{C^t_1, ..., C^t_K\}$ be the set of detected cells in this frame. These are used to initialise the tracker at step $t$. For cell $C^t_i$, the predicted locations by the tracker in $I_{t+1}$ and $I_{t-1}$ are referred as forward ($F^t_i$) and backward ($B^t_i$) predictions, respectively. Collision and mitosis are then detected, descriptions of which follow below. Note that the movement prediction model explained here does not depend on the morphology of the cell or on any other image property. Hence, it is directly applicable on top of any segmentation algorithm without the need for additional tuning.

**Collision detection.** Collision occurs when two cells share a fraction of their boundary, and this can be mistaken as a single cell during segmentation. When processing a new frame $I_t$, where $t > 1$, collision detection is performed first in which a cell $C^t_i$ is considered to be a lump of multiple individual cells if the centroids of two or more cells in $S_{t-1}$ lie within the tracked region $B^t_i$. If this is the case, $C^t_i$ is re-segmented using the centroids of the two cells in the previous frame $I_{t-1}$. This collision detection procedure continues until each cell in $I_t$ matches at most one cell in $I_{t-1}$.

**Mitosis detection.** Henceforth, a similar procedure as for the collision detection is performed for mitosis but in the opposite direction for $t$ in the sequence of frames. Cells are matched in $S_{t-1}$ to the detected cells in $I_t$. Namely a cell $C^t_{i-1}$ is matched to a cell $C^t_i$ if the centroid of $C^t_i$ is inside the region $F^t_{i-1}$. Different from collision detection, however, $C^t_{i-1}$ is also matched to $C^t_i$ if the centroid of the region $F^t_{i-1}$ lies within the boundaries of the cell $C^t_i$. This matching procedure yields a set of matches for each cell $C^t_{i-1}$, which we denote as $M^t_{i-1}$ and its size as $|M^t_{i-1}|$. The state of cell $C^t_{i-1}$ is then determined according to Equation (1).

$$C^t_{i-1} = \begin{cases} \text{Apoptosis}, & |M^t_{i-1}| = 0 \\ \text{Mitosis}, & |M^t_{i-1}| = 1 \\ \text{Mitosis}, & \text{otherwise} \end{cases} \quad (1)$$

In case of mitosis, the cell splits, thus the tracking of $C^t_{i-1}$ ends and the cells in $|M^t_{i-1}|$ are initialised with two new trackers which have $C^t_{i-1}$ as their parent. The new cells in $S_t$
that are not linked to any cell in $I_{t-1}$ are interpreted as newly detected cells which start their “life” in $I_t$ without link to a parent cell.

**Re-segmentation.** In case of a detected collision of two or more cells into a cell $C_i^t$, cell $C_i^t$ is re-segmented in such a manner that the new number of segments matches the number of colliding cells. This is achieved using watershed deconvolution as described in the work of Kachouie et al. (2008). To prevent over-segmentation of the cell $C_i^t$, which adversely affects segmentation accuracy, the relative position of the centroids of the cells in $I_{t-1}$ are used as the seeds for the segmentation algorithm. An illustration of the re-segmentation of cells is shown in Figure 3 in Appendix C.

The above algorithm is designed as a simple and general method to track biological cells of different size, shape and movement patterns. Volatile trajectories and unpredictability of cell location are dealt with using one approach for all datasets with the re-segmentation being invariant to cell morphology and image properties.

### 3. Experiments

#### 3.1. Data

The effectiveness of the proposed methodology is demonstrated through applications on three datasets listed on ISBI (Maška et al., 2014), namely PhC-C2DL-PSC, Fluo-N2DH-SIM+, and DIC-C2DH-HeLa. The first two datasets contain stem cells and nuclei of cells, respectively, shaped as small ellipsoids. The PhC-C2DL-PSC dataset contains significantly larger number of cells and an active population of cells moves very erratically (Maška et al., 2014). For Fluo-N2DH-SIM+, the cell movements are moderate, and for DIC-C2DH-HeLa, they are minimum. For DIC-C2DH-HeLa, the cell morphology deviates from a round circle and lacks a well defined perimeter boundary due to small cell structures on the edge of the cells. This characteristic makes the cell segmentation more difficult when combined with the other datasets to be segmented by one single approach. Two sequences per dataset are available for training, and another set of two sequences per dataset is used for testing (Maška et al., 2014).

#### 3.2. Results

The performance of our methods is evaluated using the same measures as described in Maška et al. (2014). For segmentation, the average of detection (DET) and segmentation (SEG) metrics, expressed as $OP_{CSB} = \frac{1}{2}(DET+SEG)$, is used. Further, for tracking, the average of SEG and tracking (TRA) measures, expressed as $OP_{CTB} = \frac{1}{2}(SEG+TRA)$, is used.

Table 1 provides the performance scores for the three variants of TAS. As expected, the performance of **TAS-specialised** is the best, while that of **TAS-general** is the lowest. Further, to validate our **TAS-general** method, we compare it with a baseline method. For baseline, we choose the results of Zhou et al. (2019), since this approach as well requires minimal parameter adjustments. The comparison on Fluo-N2DH-SIM+ dataset shows that our general approach outperforms the baseline. The intermediate semi-general, **TAS-intermediate** variant further improves the performance, mainly due to the more precise boundary refining of ellipsoid shapes.
We further compare the performance of our approach with the state-of-art methods. We compare with the published works of Zhou et al. (2019) and Lux and Matula (2019), and the top 3 submissions on the leaderboard of ISBI cell tracking challenge. These approaches are highly tuned to the type of cell morphology and image properties, hence we implement a \textit{TAS-specialised} variant for comparison. Table 2 presents the results of our method compared to the other approaches. The detailed results, including all metrics, are shown in Table 4 in Appendix A. As can be seen, our approach outperforms all other methods for datasets PhC-C2DL-PSC and Fluo-N2DH-SIM+, and ranks second for the DIC-C2DH-HeLa.

An interesting observation is that on the tracking metric (TRA) for the DIC-C2DH-HeLa dataset, our approach still performs the best with a value of 0.955, while the previous first place approach scores slightly lower. The cells in DIC-C2DH-HeLa lack easily discernible boundaries compared to other datasets, and lack intense movement activities. This is primarily the reason that this dataset does not benefit from our modelling of cell behaviour. The minor shift of the TRA metric can be explained by the explicit collision detection method which the method of Lux and Matula (2019) lacks. In datasets Fluo-N2DH-SIM+ and PhC-C2DL-PSC, highly active cells are better detected by our tracking method. This is clearly evident in PhC-C2DL-PSC, where up to 1000 cells are present in each frame, compared to a maximum of around 70 in the Fluo-N2DH-SIM+. Collisions and mitotic events are more evident with increased spatial displacement which constitutes our method an ideal tracker and re-segmentation. Due to this reason, our method outperforms the previous state-of-art-approaches by a larger margin on this dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>(\text{OP}_{\text{CSB}})</th>
<th>(\text{OP}_{\text{CTB}})</th>
<th>(\text{OP}_{\text{CSB}})</th>
<th>(\text{OP}_{\text{CTB}})</th>
<th>(\text{OP}_{\text{CSB}})</th>
<th>(\text{OP}_{\text{CTB}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhou et al. (2019)</td>
<td>0.861</td>
<td>0.860</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{TAS-general}</td>
<td>0.872</td>
<td>0.870</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{TAS-intermediate}</td>
<td>0.895</td>
<td>0.893</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{TAS-specialised}</td>
<td>\textbf{0.897}</td>
<td>\textbf{0.896}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: \(\text{OP}_{\text{CSB}}\) and \(\text{OP}_{\text{CTB}}\) results of the three different initial segmentation approaches, as described in section 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>DIC-C2DH-HeLa</th>
<th>Fluo-N2DH-SIM+</th>
<th>PhC-C2DL-PSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhou et al. (2019)</td>
<td>-</td>
<td>0.861</td>
<td>0.806</td>
</tr>
<tr>
<td>Lux and Matula (2019)</td>
<td>0.894</td>
<td>-</td>
<td>0.809</td>
</tr>
<tr>
<td>ISBI CTC\textsuperscript{1} 3rd entry</td>
<td>0.884</td>
<td>0.887</td>
<td>0.808</td>
</tr>
<tr>
<td>ISBI CTC\textsuperscript{1} 2nd entry</td>
<td>0.895</td>
<td>0.890</td>
<td>0.809</td>
</tr>
<tr>
<td>ISBI CTC\textsuperscript{1} 1st entry</td>
<td>0.912</td>
<td>0.896</td>
<td>0.814</td>
</tr>
<tr>
<td>\textit{TAS-specialised}</td>
<td>0.905</td>
<td>\textbf{0.897}</td>
<td>\textbf{0.846}</td>
</tr>
</tbody>
</table>

Table 2: \(\text{OP}_{\text{CSB}}\) and \(\text{OP}_{\text{CTB}}\) scores, as of 30th of January for the entries to the ISBI Cell Tracking Competition (CTC)\textsuperscript{1} (Máška et al., 2014). MU-Lux-CZ the same team as in the work of Lux and Matula (2019), ND-US, BGU-IL the same team as in the work of Zhou et al. (2019), CVUT-CZ, HD-Hau-GE, UVA-NL.
4. Discussion

Since our approach enhances the tracking capabilities of segmentation focused methods, it is important to identify which aspects of the tracker contribute most to the improvements reported earlier. Table 3 presents an ablation study involving TAS-specialised being studied with respect to the collision and mitosis modules. As can be seen, both collision and mitosis modelling steps are an integral part of the approach, since there exists a clear dependence on them. Removing the re-segmentation correction completely results in lower performance values. However, it is important to note that the performance variation between different model variants are not very large, with a mean standard deviation value of 0.0169. This is attributed to the saturation of the performance of state-of-the-art approaches on this challenge which does not allow for a large improvement margin.

5. Conclusions

Medical images of biological cells, contain many noise artefacts, convoluted cell boundaries and unpredictable cell movements which often confuse cell segmentation methods. We propose a siamese tracking assisted re-segmentation approach which specifically models biological cell activities (mitosis, apoptosis and cell collisions) and enhances cell segmentation. Our results indicate that biologically inspired tracking models of micrometer-scaled cells can better apprehend erratic behaviour. We demonstrate the applicability of our method on three cell tracking datasets. The specialised variant outperforms the previous state-of-the-art models, and ranks first on two and second on one of the three benchmark datasets.
Acknowledgments

Acknowledgments withheld.

References


### Appendix A. Additional Results

<table>
<thead>
<tr>
<th>Competition rank</th>
<th>DIC-C2DH-HeLa</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd entry</td>
<td>0.948</td>
<td>0.820</td>
<td>0.909</td>
<td>0.884</td>
<td>0.848</td>
</tr>
<tr>
<td>2nd entry</td>
<td>0.979</td>
<td>0.807</td>
<td>0.969</td>
<td>0.887</td>
<td>0.882</td>
</tr>
<tr>
<td>1st entry</td>
<td>0.960</td>
<td>0.665</td>
<td>0.950</td>
<td>0.808</td>
<td>0.804</td>
</tr>
<tr>
<td><strong>TAS-specialised</strong></td>
<td>0.958</td>
<td>0.852</td>
<td><strong>0.955</strong></td>
<td>0.905</td>
<td>0.904</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competition rank</th>
<th>Fluo-N2DH-SIM+</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd entry</td>
<td>0.956</td>
<td>0.834</td>
<td>0.954</td>
<td>0.895</td>
<td>0.894</td>
</tr>
<tr>
<td>2nd entry</td>
<td>0.981</td>
<td>0.813</td>
<td>0.973</td>
<td>0.890</td>
<td>0.889</td>
</tr>
<tr>
<td>1st entry</td>
<td>0.984</td>
<td>0.682</td>
<td>0.957</td>
<td>0.809</td>
<td>0.804</td>
</tr>
<tr>
<td><strong>TAS-specialised</strong></td>
<td>0.972</td>
<td><strong>0.822</strong></td>
<td>0.971</td>
<td><strong>0.897</strong></td>
<td><strong>0.896</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competition rank</th>
<th>PhC-C2DL-PSC</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd entry</td>
<td>0.961</td>
<td>0.863</td>
<td>0.954</td>
<td>0.912</td>
<td>0.909</td>
</tr>
<tr>
<td>2nd entry</td>
<td>0.983</td>
<td>0.821</td>
<td>0.975</td>
<td>0.896</td>
<td>0.895</td>
</tr>
<tr>
<td>1st entry</td>
<td>0.967</td>
<td>0.715</td>
<td>0.959</td>
<td>0.841</td>
<td>0.836</td>
</tr>
<tr>
<td><strong>TAS-specialised</strong></td>
<td><strong>0.972</strong></td>
<td><strong>0.720</strong></td>
<td><strong>0.966</strong></td>
<td><strong>0.846</strong></td>
<td><strong>0.843</strong></td>
</tr>
</tbody>
</table>

Table 4: Detailed performance comparison of our approach with the top 3 performers on the leaderboard of the cell tracking challenge, as of 30th of January. The different color codes correspond to different teams, namely MU-Lux-CZ (as in Lux and Matula (2019)), ND-US, BGU-IL (as in Zhou et al. (2019)), CVUT-CZ, HD-Hau-GE, UVA-NL, HIT-CN, FR-Ro-GE, KTH-SE and TUG-AT.

### Appendix B. Implementation details

The step by step approach to construct the initial segmentation of **TAS-general** is as follows. Raw images are normalised first by equally distributing 256 bins of pixel values, using the cumulative distribution of its histogram. To enhance the contrast between maximum bright spots, cell bodies, and background information in the image, the logarithm function is applied before the histogram equalisation procedure. This will relate areas closer to cell bodies and help determine more stable boundaries due to minimised effect of standalone maximum values. Next, the image values are projected to the range of 0-1, and then normalised further to zero mean and unit variance. This pre-processing step is applied the same way to all images of any dataset and fed into the U-Net network to produce cell body predictions.

Thereafter, the Euclidean distance from the background is used as a successive threshold metric. What this entails is acquiring regions where the distance is above 0.5, if 0 indicates background information and 1 is the farthest area from it. Thereafter, these regions are compared with the bigger regions cutoff at 0.1. If there exists a region that was not identified by the stricter cutoff value of 0.5, then the new region is considered to be a cell centroid. This
is performed again for a cutoff value of 0, which coincides with the exact U-Net prediction. The last step is necessary to detect small cells appearing in the frame, which the cutoff values 0.1 and 0.5 would ignore. An additional condition for a cell centroid is that is at least bigger than the third of the mean cell size detected at each step. This condition is necessary to reject small outlier pixels. The final cell centroid regions are used as seeds for the random walker approach to define the boundary of each smaller cell contained in the initial cell body prediction of the U-Net.

Appendix C. Additional Illustrations

Figure 3: Schematic representation of cell collision detection using a Siamese tracker, and re-segmentation of the detected cells using watershed approach. The two cells, far apart in \((t-2)^{th}\) frame, collide in the \(t^{th}\) frame, and are wrongly segmented as a single cell. Through siamese tracking between \((t-2)^{th}\), \((t-1)^{th}\) and \(t^{th}\) frames, the collision event is identified, and applying watershed approach over the \(t^{th}\) frame helps to correct the segmentation.

Figure 4: Biological cell movement behaviours between subsequent frames.