Medical Image Analysis 77 (2022) 102358

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

ACME: Automatic feature extraction for cell migration examination through intravital microscopy imaging

Miguel Molina-Moreno[a,*], Iván González-Díaz[a], Jon Sicilia[b], Georgiana Crainiciuc[b], Miguel Palomino-Segura[b], Andrés Hidalgo[b], Fernando Díaz-de-María[a]

[a] Universidad Carlos III de Madrid, Avda. de la Universidad 30, Leganés 28911, Spain
[b] Area of Cell and Developmental Biology, Centro Nacional de Investigaciones Cardiovasculares Carlos III, Melchor Fernández Almagro 3, Madrid 28029, Spain

ARTICLE INFO

Article history:
Received 23 June 2021
Revised 7 January 2022
Accepted 9 January 2022
Available online 14 January 2022

MSC:
68T07
92B20

Keywords:
3D Segmentation
Cell tracking
Convolutional neural networks
Cell migration

ABSTRACT

Cell detection and tracking applied to *in vivo* fluorescence microscopy has become an essential tool in biomedicine to characterize 4D (3D space plus time) biological processes at the cellular level. Traditional approaches to cell motion analysis by microscopy imaging, although based on automatic frameworks, still require manual supervision at some points of the system. Hence, when dealing with a large amount of data, the analysis becomes incredibly time-consuming and typically yields poor biological information. In this paper, we propose a fully-automated system for segmentation, tracking and feature extraction of migrating cells within blood vessels in 4D microscopy imaging. Our system consists of a robust 3D convolutional neural network (CNN) for joint blood vessel and cell segmentation, a 3D tracking module with collision handling, and a novel method for feature extraction, which takes into account the particular geometry in the cell-vessel arrangement. Experiments on a large 4D intravital microscopy dataset show that the proposed system achieves a significantly better performance than the state-of-the-art tools for cell segmentation and tracking. Furthermore, we have designed an analytical method of cell behaviors based on the automatically extracted features, which supports the hypotheses related to leukocyte migration posed by expert biologists. This is the first time that such a comprehensive automatic analysis of immune cell migration has been performed, where the total population under study reaches hundreds of neutrophils and thousands of time instances.

1. Introduction

Microscopic imaging modalities are widely used thanks to their versatility and moderate cost (Dhawan et al., 2010). In particular, confocal and epifluorescence microscopy imaging is the most popular method for studying the dynamic behavior exhibited by live cells. Fluorochromes and fluorescent proteins have enabled different cells and cellular components to be identified with a high specificity in the context of nonfluorescing material, and thus, have facilitated the visualization and determination of the relationships and interactions between them (Combs and Shroff, 2017). However, fluorescence microscopy imaging has significant limitations: temporal decay of fluorochromes capacity owing to photobleaching, phototoxicity, and poor resolution on account of the size of the structures being captured (on the order of micrometers) (Jensen, 2012). Consequently, both data acquisition and data analysis can be challenging, and typically low through-put.

Cell segmentation and tracking in multidimensional microscopy imaging has arisen as one of the fundamental needs for the biomedical community. For example, recent studies have revealed the relationships between infarcts and the deformation and motility of neutrophils along the blood vessels (García-Prieto et al., 2017), or between sickle cell anaemia and heterogeneous cell adhesion and deformability of red blood cells (Barabino et al., 2010). Due to the specifics of fluorescence microscopy imaging data, their analysis poses three major challenges: 1) given the increasingly larger volumes of available data, manual or semi-automated segmentation and tracking have become unfeasible in most applications (Orth et al., 2017), especially for multidimensional imaging; 2) fluorescence presents high intra- and inter-plane variability, and microscopic imaging exhibits poor resolution, hampering cell segmentation; and 3) cells move freely along the volume following erratic paths, which collide and separate in multiple points. Thus,

* Corresponding author.
E-mail address: mmolina@tsc.uc3m.es (M. Molina-Moreno).

https://doi.org/10.1016/j.media.2022.102358

a robust segmentation system has to be able to adapt to different contrast, illumination and resolution conditions, and an effective tracking module has to be able to handle cell collisions and separations.

Notwithstanding the microscopic imaging potential, automatic analysis of microscopy images is yet an emerging subject of research, especially when dealing with 3D volumes. Although Convolutional Neural Networks (CNNs) have shown impressive performance in many computer vision tasks over real images, such as classification (Krizhevsky et al., 2012), segmentation (Shelhamer et al., 2017) and object detection (Ren et al., 2015), the interpretation of 3D medical images requires significant expert knowledge, which is often difficult to embed in a CNN architecture. Furthermore, the lack of large and annotated datasets and the substantial differences between the distinct imaging modalities hinder the use of general purpose CNNs. In consequence, current systems are typically application-dependent (Chen et al., 2019a).

Despite the research efforts devoted to 3D segmentation and tracking to overcome these obstacles, biologists have generally addressed the problem of cell segmentation and tracking using commercial software, such as Imaris software (Bitplane, South Windsor, CT, USA) or Fiji (Schindelin et al., 2012). These libraries, although implement some standard algorithms for image segmentation and tracking, require manual supervision as they are not robust enough to tackle all potential scenarios and data. In addition, previous attempts to automatize cell migration analysis (involving 3D CNNs and fluorescence microscopy imaging) required manual supervision in 3D segmentation (Lefevre et al., 2020; Henry et al., 2013) and 4D cell tracking (Richards et al., 2018). Manual supervision and filtering is therefore a fundamental step in current solutions to ensure that only cells which are correctly segmented and tracked are considered in the subsequent analysis and, thus, to avoid potential biases due to the inclusion of non-relevant regions, artifacts and noise. A straight consequence is that biological studies have been restricted to tens to few hundred of cells (Pijuan et al., 2019), as in our previous studies García-Prieto et al. (2017) and Sreeramkumar et al. (2014). Moreover, the state-of-the-art approaches concerning 3D segmentation do not tackle the task of 4D tracking (Jaeger et al., 2018; Tokuoka et al., 2020).

Here, we introduce ACME (Automatic Cell Migration Examination), an automatic feature extraction method for cell migration analysis within blood vessels in microscopy imaging. To the best of our knowledge, this is the first fully-automatic system that implements the whole processing pipeline for 3D cell segmentation and tracking of 4D microscopic imaging. By combining deep learning and machine learning blocks, we have built a pipeline capable of segmenting, tracking and extracting features from cells moving within blood vessels. Furthermore, an explainability model automatically discovers cell behaviors and profiles them in terms of the extracted features. In our case of study, a blood vessel contains a series of neutrophils, which, in the context of inflammation, migrate and dynamically adapt their shape to efficiently moving towards the lesion area. However, ACME it is versatile and could be adapted to different scenarios, such as the migration of immune cells in tumor samples (Di Pilato et al., 2021a).

As we will demonstrate in the experimental section, our approach automatically segments and tracks the cells in the 4D (3D space plus time) volumes, reducing drastically the human effort devoted to the annotation (i.e. the time required) and removing human biases in cell selection (e.g. choosing cells that are simpler to analyze). Moreover, our system increases the throughput of analyzed data by more than an order of magnitude: in our dataset, by setting a very high system precision that ensures that only accurate segmentations and tracks are considered, ACME is able to automatically extract features from thousands of temporal instances of hundreds of cells, instead of tens of cells, the limit in our previous studies based on manual analysis (García-Prieto et al., 2017; Sreeramkumar et al., 2014). In addition, the shape- and motion-based feature extraction (we propose an exhaustive set of features inspired by cell dynamics) allows ACME to automatically provide some insights into the observed cell behaviors, enhancing our understanding of complex physiological processes at microscopic and single cell resolution, and leading to observations and conclusions that are in agreement with our experimental-based hypotheses.

Specifically, the main contributions of this work can be summarized as follows:

- The most significant scientific contribution in the paper is that ACME incorporates a strongly-regularized 3D CNN which, being trained with pixel-wise annotations, is able to segment jointly blood vessels and cells.
- In addition, we have designed a robust multi-target tracking system that is able to detect and handle collisions between cells, keeping their individual tracks.
- ACME includes a novel and comprehensive feature extraction method that, relying on two coordinate systems adapted to the geometry of the vessel (polar and cartesian aligned with vessel direction), characterizes the short-term behavior of cells through the analysis of their locations, shapes, and motion relative to the blood vessel.
- We introduce a high-precision cell selection module that, while keeping enough available cells for subsequent processing, guarantees that the analysis of extracted features is not biased by the inclusion of wrongly detected non-cell regions.
- Finally, we have designed a hierarchical explainability model which automatically discovers the most prominent cell behaviors and relates them to meaningful subsets of relevant features, enabling the researchers to gain insight into biological processes related to cell motion.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. In Section 3 we first provide a general description of our method of automatic feature extraction for the analysis of cell migration in microscopy imaging, and then describe in detail of each constituent processing block. Section 4 explains our proposed hierarchical model that automatically discovers prominent cell behaviours and provides meaningful insights, thus increasing explainability. Section 5 describes and discusses the experimental results that support our method whereas, finally, Section 6 summarizes our conclusions and outlines future lines of research.

2. Related work

In this section, we discuss automatic 3D segmentation methods found in the literature as well as some works that, similarly to ACME, combine 3D segmentation and tracking.

Concerning 3D segmentation methods, we start by reviewing more traditional approaches, which can be broadly divided into two groups that frequently intersect: a) those based on morphological operations and histogram analysis; and b) those relying on energy minimizing algorithms. Regarding the first group, Cheng et al. proposed a simple 3D cell segmentation method based on an adaptive thresholding followed by a splitting algorithm which relied on the size of the segmented regions (Cheng et al., 2017). In (Padfield et al., 2008), the 3D images were concurrently denoised and segmented by combining wavelet coefficients at various levels, thus enabling the extraction of cells in images with low contrast-to-noise ratios. With respect to the second group, in (Dufour et al., 2011), 3D active meshes were proposed. 3D active meshes are discrete deformable models which represent a surface as a triangular mesh and minimize an energy functional based on image intensity, gradient magnitude and the geodesic length of the contour

as a regularization term. Pécot et al. suggested a Conditional Random Field (CRF) to model the segmentation of vesicles and time-varying background estimation at the cellular scale (Pécot et al., 2015). This process performs an energy minimization using a min cut-max flow algorithm. Energy minimization methods were more effective at dealing with to fluorescence imaging challenges, like photobleaching, than the morphology-based ones.

In addition, other methods combined both techniques into hybrid approaches. In (Harder et al., 2011), a pre-processing using morphological operations, adaptive histogram thresholding and an active contour evolving the initial contours, were combined. Kong et al. suggested a segmentation method based on the gradient vector flow and an energy minimization derived from the original edge map, followed by an adaptive histogram thresholding to accurately segment each cell instance (Kong et al., 2015). In general, traditional approaches are useful when contrast is enough to visualize cell edges and cells do not collide. However, in more demanding and cluttered scenarios such as the one addressed in this paper, their performance decays considerably.

Additionally, in the last few years, several CNN-based methods for 3D segmentation of medical imaging have emerged. V-Net (Milletari et al., 2016) and, especially, the 3D version of U-Net (Çiçek et al., 2016) were specifically designed and have been widely adopted in medical imaging segmentation problems. More recently, other approaches have delved into this topic introducing more complex architectures to improve the results for particular tasks. For example, in (Chen et al., 2019b) they proposed a hybrid segmentation network combining a lightweight 3D CNN to learn long-range 3D contextual information with a 2D CNN to learn fine-grained semantic information, which has proved to be efficient in lung cancer segmentation in CT scans. In (Kamnitsas et al., 2016), a dual-pathway CNN was proposed to segment brain lesions in MRI scans. It combines a multi-scale 3D CNN which analyzes the image at two different resolutions and a fully connected 3D CRF for post-processing. However, MRI and CT scans are high resolution imaging techniques, as opposed to microscopy imaging.

Another research trend has tried to address detection and segmentation tasks using the same unified network. In (Jaeger et al., 2018), Retina U-Net was proposed as a hybrid segmentation and detection CNN, where coarser levels were used for object detection (with classification and regression blocks), and high-resolution levels were used for the auxiliary task of segmentation over the whole volume. A more recent approach (Tokuoka et al., 2020) introduced a 3D CNN to extract several quantitative criteria of embryogenesis in fluorescence imaging, including both segmentation and detection sub-networks. Its reported performance (in terms of IoU) is above 0.7, which is a good performance level for 3D segmentation in the task of cell instance segmentation during embryogenesis.

In addition to these segmentation methods, some systems have been proposed that jointly tackle the tasks of 3D segmentation and tracking. Henry et al. proposed a semi-automatic framework for the study of neutrophil migration (Henry et al., 2013). The system carried out an user modifiable Otsu's thresholding segmentation, tracked the neutrophils with a keyhole model which linked neighbouring neutrophils with similar directions in consecutive time instants, and extracted a series of features over the tracked neutrophils. Additionally, a fully automated system was proposed in (Fazli et al., 2018), consisting of a pre-processing module for MRI brain scans, cell detection based on histogram-based thresholding and tracking with the Hungarian algorithm for point matching. These systems were based on simple segmentation algorithms that are not robust enough to tackle difficult scenarios. In fact, they were tested with only 9 and 2 4D volumes, respectively. Finally, a system was proposed in (Lefevre et al., 2020), with a semi-automated segmentation module based on Trainable Weka 3D software (Arganda-Carreras et al., 2017) (an ImageJ plugin), and a tracking method with collision detection and separation relying on a watershed algorithm. In this case, manual segmentation and the ImageJ interactive environment can become a bottleneck with large-scale datasets.

In this paper, we propose a fully automated framework to characterize cell migration in microscopy imaging, called ACME. Although others (Henry et al., 2013) already modeled neutrophil motion towards a known lesion area in zebrafish, we propose a significantly enhanced characterization, based on both shape- and motion-based features computed and referenced with respect to the blood vessel surface. To the best of our knowledge, this is the first attempt to model the complete cell migration processes within blood vessels without any constraint, and to perform an automatic behavioral profiling at the cellular level. For that purpose, we have developed a robust pipeline that performs an efficient, unbiased and unsupervised analysis of cell migration, which allows the biological community to analyze the spatio-temporal behavior of cells.

3. ACME: Automatic cell migration examination

In this section we first give a general description of our fully automated system; then, a detailed explanation of its constituent processing blocks will be provided in subsequent subsections.

The full pipeline of ACME, depicted in Fig. 1, can be broken down into four modules:

1. Our system receives a 4D volume as input, $\mathbf{B} \in \mathbb{R}^4$, with 3 spatial dimensions (x, y, z) and time t. The volume is decomposed into T temporal steps, giving place to 3D volumes $\mathbf{B}_t,\ t \in [1, T]$, and each one is fed into the *3D joint segmentation module*. This module generates three outcomes: 1) a mask $\mathbf{V}_t$ defining the region of the 3D block corresponding to the blood vessel; 2) a set of N_t binary masks $\{\mathbf{C}_t\}$, containing the regions candidate to be cells $c_t^n,\ n \in [0, N_t]$; and, c) for each candidate region n, its probability of being a real cell, p_t^n. Let us note that the number of binary mask N_t may vary among frames in a sequence, as some cells may enter into or leave the region of the blood vessel being captured. A detailed description of this block will be given in Section 3.1.
2. Next, a 4D cell volume, $\mathbf{C}$, is formed concatenating the sequence of segmented 3D regions $\mathbf{C}_t$ and passed to the *three-pass 3D tracking module*. This system analyzes the time sequence of segmented regions and generates the trajectories $r_i,\ i \in [0, I]$, where I is the total number of trajectories identified in the 4D volume $\mathbf{B}$. In addition, the tracking system further refines and updates the input 4D volume $\mathbf{C}$, by handling collisions between cells. This tracking system consists of three sub-modules: 1) 3D Kalman filter tracking (Welch and Bishop, 1995), 2) collision detection, and 3) morphological post-processing, which will be described in Section 3.2.
3. Given the set of trajectories, $\mathbf{R}$, together with the blood vessel $\mathbf{V}$ and cell segmentations $\{\mathbf{C}\}$, the *feature extraction module* is in charge of characterizing the cell dynamics within short-time windows. Specifically, short-term features $\mathbf{F_R}$ are formed by aggregating two types of features: a) instantaneous features, related to the position and shape of the cells in each time instant; and b) dynamic features, representing both the evolution of these instantaneous variables in a short-time window and other properties defining the cell trajectory. More details about the features extraction module are given in Section 3.3.
4. The *cell selection module* is the last step of the analysis system. Relying on the short-term features and the previously computed probabilities p_t^n, and applying some filters that enforce certain constraints in accordance with the biological character-

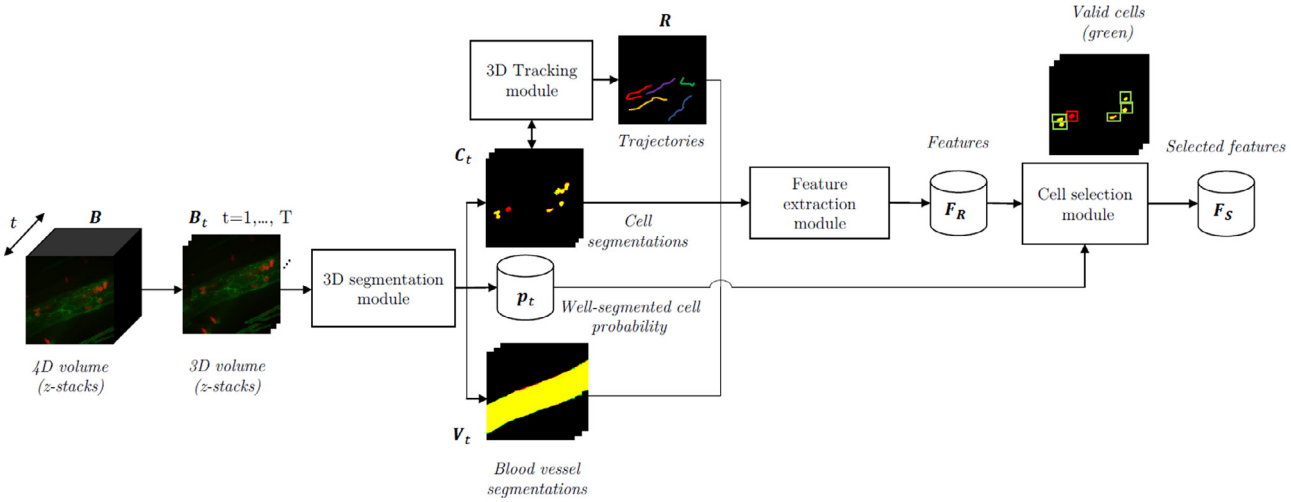


Fig. 1. Overview of the processing pipeline of ACME. Each 3D block corresponding to a time instant of a 4D volume is fed into the 3D joint segmentation module, which produces a blood vessel 3D mask, a set of individual cell 3D masks and their well-segmented probabilities (indicating, for each cell, its probability of being well-segmented). Then, considering the 3D cell data over time, the 3D tracking method proposes the trajectory of each potential cell. Next, the feature extraction module analyzes each potential cell and its trajectory and to provide a comprehensive set of instantaneous and dynamic features. Finally, the cell selection module, according to the features and the well-segmented probabilities, discards those segmented regions not likely to be a cell.

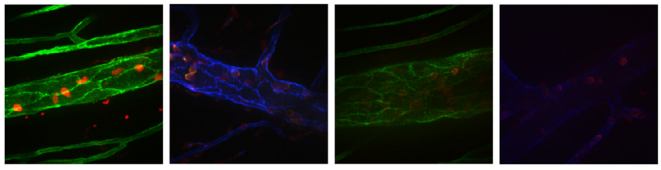


Fig. 2. Illustrative examples of 3D time instances of 4D volumes of intravital fluorescence microscopy (z-stacked). Samples are heterogeneous in terms of resolution, luminiscence and blurring, especially in cell channels.

istics of the cells, this module selects, from the entire collection **R**, the final set of valid trajectories $\mathbf{S} = s_j,\ j \in [0, J]$, where J is the total number of valid trajectories. A more detailed explanation is given in Section 3.4.

Finally, we build a feature database including the short-term features $\mathbf{F_S}$ associated with every time instant t of the selected trajectories **S**. Once the feature database is created, the different cell behaviors can be explained from the features, as described in Section 4.

3.1. 3D Strongly-Regularized segmentation module

Although there exist networks tailored to perform semantic segmentation in medical imaging (Çiçek et al., 2016; Jaeger et al., 2018), *in vivo* cell microscopic imaging within blood vessels poses particular challenges; particularly:

1. Cell pixel proportion is small with respect to the other two classes (vessel and background pixels), leading to a heavy class-imbalance problem that should be properly addressed. Let us note that we consider the full vessel volume, and not only its walls, as fluorochromes affect the full volume.
2. Intravital fluorescence microscopic imaging methods are unstable: they are heavily affected by subject reaction to contrast agents and movement, photobleaching, poor resolution and blurring. Hence, the data extracted are variable in terms of luminiscence, noise and contrast, as illustrated in Fig. 2.
3. For the specific scenario tackled in this paper, maximizing cell detection precision (i.e., minimizing false positives) is essential: the system must properly detect the cells with at least 95% precision to ensure that the subsequent cell behavior analysis is not biased by the presence of badly segmented cells or artifacts (although this requirement might be relaxed according to the system robustness).

The 3D strongly-regularized segmentation CNN architecture is depicted in Fig. 3. The regularization process aims to improve the cell and blood vessel segmentation by the use of multiple CNN branches devoted to: link blood vessel and cell segmentations, improve the robustness and maximize cell detection precision. Specifically, we have built three subsystems on top of a 3D U-Net backbone (Çiçek et al., 2016) to address each one of the challenges described above; namely:

1. *Addressing class-imbalance*: Since the segmentation quality benefits from addressing the segmentation of both the blood vessel and the cells using a single 3D CNN, we need to tackle the aforementioned class-imbalance problem. In particular, we have designed a two-branch segmentation head (see Fig. 3). The first branch ends with a conventional multi-class cross entropy, L_{CE}, considering the original 3-class problem (vessel, cells and background pixels); whereas the second branch considers a binary cell-vs-rest problem and enforces the CNN to more pay attention to the cell segmentation. To that end, we have introduced a regularization loss in the form of α-balanced focal loss (Lin et al., 2017). Working on a binary problem has led to better results than simply regularizing the multi-class branch. The rationale behind is that the joint imbalance between the three classes made especially difficult to adjust the corresponding weights for both vessel and cells.
To be more precise, let us consider a cell-vs-rest binary segmentation problem, where the label $y = 1$ means that a pixel belongs to a cell and $y = 0$ that is either background or vessel, and define $p \in [0, 1]$ as the model's estimated pixel probability for the cell class. Hence, the α-balanced focal loss is:

$$FL_{BCE} = -\alpha(1 - \tilde{p})^{\gamma} \log(\tilde{p}) \tag{1}$$

where $\gamma \geq 0$ is a tunable focusing parameter, α is the cell-vessel sample proportion factor, and a $\tilde{p}$ is the model's estimated probability for the ground-truth class y of each pixel ($\tilde{p} = p$ if $y = 1$, and $\tilde{p} = 1 - p$, otherwise). Apart from this term, we add a standard Dice Loss (Milletari et al., 2016), as seen in Eq. 2, to regularize the segmentation at the 3D volume level,

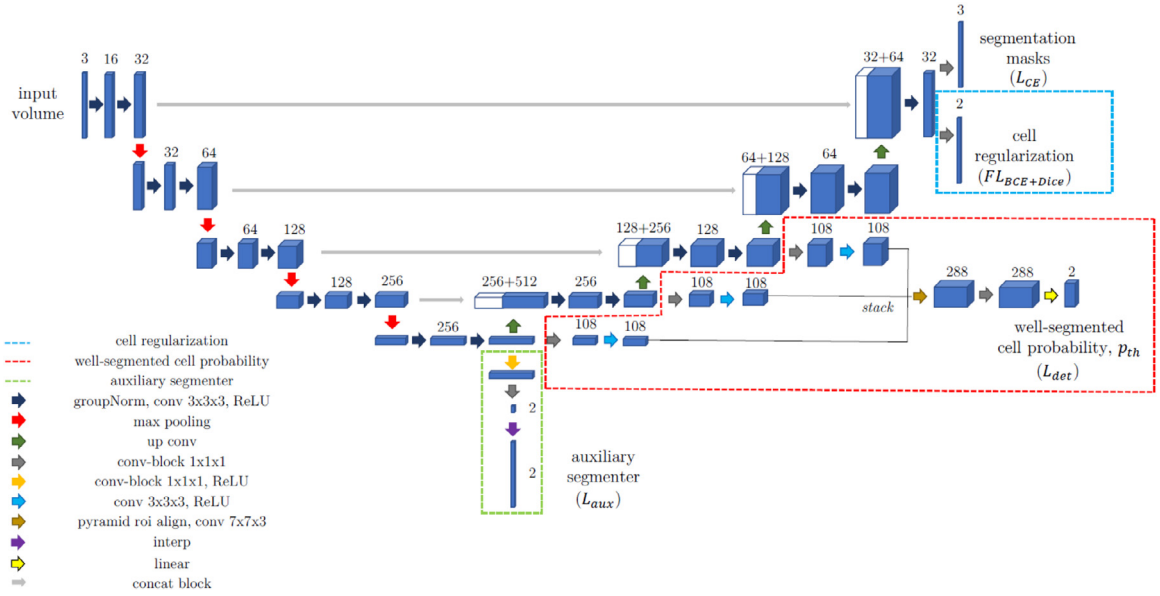


Fig. 3. 3D strongly-regularized segmentation network architecture. On top of the 3D U-Net backbone, we have built three modules: a cell regularization scheme to tackle the cell-vessel sample imbalance; an auxiliary segmenter that aims to improve the effectiveness of the training procedure on the U-Net analysis-synthesis structure; and a well-segmented probability estimation branch which allows us to improve the precision at system level.

resulting $FL_{BCE+Dice}$, with both terms equally weighted.

$$L_{Dice} = 1 - \frac{2(\mathbf{C}_t \cap \mathbf{G}_t)}{\mathbf{C}_t \cup \mathbf{G}_t}, \tag{2}$$

where $\mathbf{G}_t$ is the ground-truth binary mask corresponding to the time instant t, and the operators $\cap$ and $\cup$ the intersection and union, respectively, of the binary masks.

2. *Addressing strong data variability*: It is known that the analysis-synthesis structure of 3D U-Net hinders the gradient back-propagation in scenarios with few and variable data (Glorot and Bengio, 2010). To address this problem, we have added an *auxiliary segmenter* at the end of the analysis path (Szegedy et al., 2015), which: 1) works on a low resolution representation of the image, thus favoring the robustness and the network capacity to identify the cells in the analysis path, instead of maximizing segmentation precision (this is performed in the synthesis path); and 2) injects additional gradients, thus improving the effectiveness of the training process and, consequently, the segmentation performance. This auxiliary segmenter focuses on the cells (the vessel is easier to segment) and uses the same α-balanced focal loss described above, here named L_{aux}.
3. *Maximizing cell-detection precision*: In demanding scenarios such as the one tackled in this paper, current 3D segmentation CNNs are far from achieving the required 95% precision level for cell detection. However, such precision value is necessary to avoid biases in the subsequent cell behaviour analysis due to the presence of badly segmented cells or artifacts. To this purpose, instead of considering an instance segmentation problem (He et al., 2017; Jaeger et al., 2018), we have estimated the probability of a region being a well-segmented cell, which is used as an input to the subsequent cell selection module to considerably raise the precision level. Specifically, this probability estimation has been built through a separate branch of the CNN, which relies on 3D segmentation bounding boxes as regions of interest (RoIs) to decide whether a cell is well segmented or not. This branch is trained through a binary cross-entropy loss function named L_{det} (refer to Eq. (3)) that considers a proposal c_t^n as good ($\beta_n = 1$) when the 3D IoU between the ground truth g_t^n and this proposal is above 0.7, and as bad ($\beta_n = 0$) when the 3D IoU is below 0.5. We also explored the combination of segmentation and object detection in the same network (Tokuoka et al., 2020), but this approach failed to converge properly in our scenario.

$$L_{det} = -(\beta_n \log(p_t^n) + (1 - \beta_n) \log(1 - p_t^n)). \tag{3}$$

The network is trained through a multi-stage pipeline: first the 3D segmentation part (through L_{CE}, $FL_{BCE+Dice}$ and L_{aux}) and then the well-segmented probability estimation branch (through L_{det}), which relies on the cell segmentation mask. Finally, it is worth describing our data augmentation, which has consisted in several random transformations including independent brightness and gamma transforms over each channel, and geometric affine transformations and flips.

3.2. A three-pass 3D cell-tracking with collision handling

Cells migrating within blood vessels often describe erratic trajectories, vary their shapes, and collide with others. Hence, a robust tracking system has to be able to deal with occasional changes of direction, keep a record of potential collisions and differentiate between contacting cells. The 3D tracking system proposed in this paper addresses these challenging tasks through three passes that combine a 3D Kalman filter with specifically tailored modules for collision detection and handling, and post-processing, which successively refine the results.

The general scheme of the 3D tracking system is depicted in Fig. 4 and comprises three passes, described next.

3.2.1. Pass 1. First Kalman pass and collision handling

The first step consists in a 3D Kalman filtering followed by a collision detection and handling algorithm. Specifically, it receives the sequence of cell segmentations provided by the previous module, $\mathbf{C}_t = \ t \in [1, T]$, and performs a tracking based on a constant acceleration Kalman filter, able to follow changes in cell velocity. As a result, it generates the associated trajectory for each cell and, when applicable, an indication of a potential collision, which is stored in a collision matrix (in which case, it would require further processing).

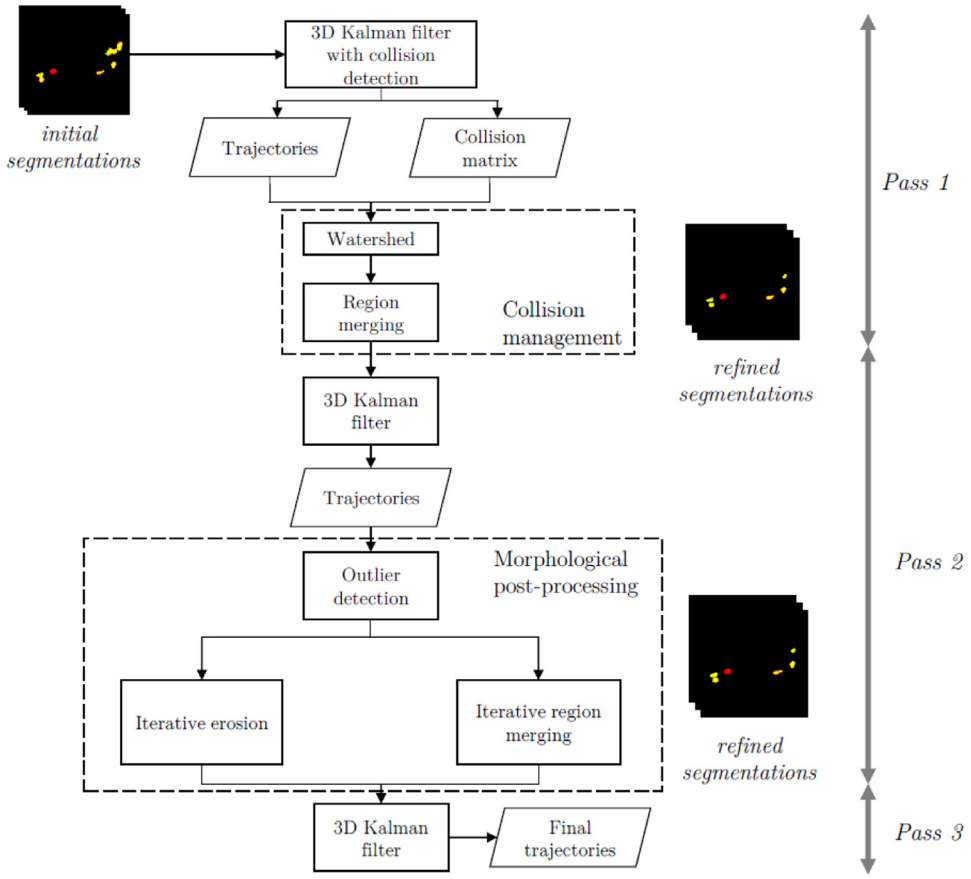


Fig. 4. Flowchart of the 3D tracking module. After the first 3D Kalman filtering process, collisions are detected and handled by reference-constrained splitting them. Next, we re-track the cells and apply a morphological post-processing on the remaining cells, depending on their relative volume with respect to the rest of instances in the trajectory. Finally, once the segmentations are modified, we perform a new tracking process that produces the final trajectories.

The Kalman Filter: A constant acceleration Kalman filter is used to model the temporal state $\mathbf{x}_t$ using a linear *state-transition model*, i.e.:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + w_t, \tag{4}$$

which describes how the system evolves from the state at time $t-1$, $\mathbf{x}_{t-1}$, to the state at time t, $\mathbf{x}_t$, with $\mathbf{A}$ being the transition matrix and w_t representing noise drawn from a zero-mean multivariate normal distribution. In our particular case, the state variable $\mathbf{x}_t$ is defined as follows:

$$\mathbf{x}_t = [\mathbf{r}_t\ \mathbf{v}_t\ \mathbf{z}_t]^T. \tag{5}$$

where $\mathbf{r}_t$, $\mathbf{v}_t$ and $\mathbf{a}_t$ are 3-dimensional vectors describing cell position, velocity and acceleration at the time instant t, respectively. Moreover, $\mathbf{A}$ takes the following form:

$$\mathbf{A} = \begin{bmatrix} I_3 & \Delta t I_3 & \frac{\Delta t}{2} I_3 \\ 0_3 & I_3 & \Delta t I_3 \\ 0_3 & 0_3 & I_3 \end{bmatrix}, \tag{6}$$

where Δt is the time step between consecutive frames ($\Delta t = 1$ for simplicity), and I_3 and 0_3 are the 3×3 identity and zero matrices, respectively.

Additionally, given the state $\mathbf{x}_t$, the tracking system generates an observation that corresponds with the location of cell $\mathbf{r}_t$ according to a linear *observation model*:

$$\mathbf{z}_t = \mathbf{O}\mathbf{x}_t + g_t \tag{7}$$

which is defined through an observation matrix $\mathbf{O}$ and an observation noise g_t, which is assumed to be a zero-mean Gaussian noise. In our particular case, the observation matrix is:

$$\mathbf{O} = \begin{bmatrix} I_3 & 0_3 & 0_3 \end{bmatrix}, \tag{8}$$

Following a multi-object tracking approach, at a given time instant t, there are two variables that need to be aligned: a) the M_t active trajectories $x_t^m,\ m \in [0, M_t]$ (spanning over previous instants); and b) the segmented cells at t, $c_t^n,\ n \in [0, N_t]$. Note that, in general $M_t \neq N_t$, which means that new trajectories may need to be created or some of them deleted.

Tracker-Cell Assignment: The assignment process is carried out in two steps. First, we perform hard assignments: when the centroid of a segmented cell c_t^n falls within a predicted active trajectory x_t^m, the cell is assigned to it ($n \to m$) and not further considered in the assignment process. Second, we perform soft assignments. In this case we require both temporal and size and shape consistency, i.e., if the distance between the cell and the predicted trajectory is small and the cell properties also fit those of previous instants in the same trajectory, we perform a soft assignment.

Collision detection: Sometimes two cells collide or separate from each other after colliding, which causes that $M_t \neq N_t$. In order to properly handle these situations, a collision detection strategy is needed. After the assignment process is finished for a time instant t, if a segmented cell $c_t^{\tilde{n}}$ remains unassigned, we check if it comes from a previous collision (i.e., there was certain overlapping between that cell $c_t^{\tilde{n}}$ and one of the cells tracked in the previous time instant c_{t-1}^n). If so, we register the corresponding event in the collision matrix, otherwise, we consider that a new cell has entered into the scene, and we simply add a new track accordingly. And conversely, if an active trajectory $x_t^{\tilde{m}}$ remains unassigned at time t, we identify its assigned cell in the previous instant $c_t^{\tilde{n}}/\tilde{n} \to \tilde{m}$,

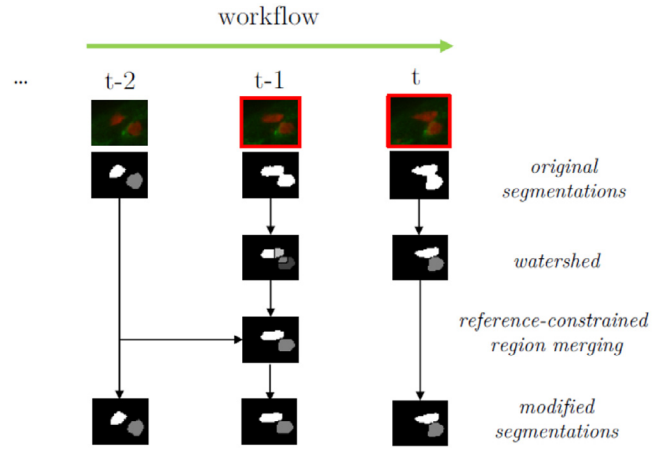


Fig. 5. 3D tracking. Illustrative example of how the splitting process works when collisions are detected. In red, time instants exhibiting detected collisions. From top to bottom, original segmentations of collided cells are processed by a watershed algorithm, whose results are refined to avoid over-segmentation by a temporal consistency-constrained region merging algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and check if it shows a certain overlapping with any of the current cells c_t^n. Again, if so, we register the corresponding event.

Splitting cell segmentations: Once collisions have been detected, a splitting algorithm is required. Current tracking systems, as the one in PhagoSight (Henry et al., 2013), use a watershed algorithm to split the detected collisions, but they do not take advantage of the temporal consistency of the cell segmentation. In our case, collisions often occur only for a few time instants. Thus, we propose to leverage temporal consistency by considering two consecutive time instants, as illustrated in Fig. 5. The proposed splitting module is also based on a watershed segmentation algorithm, but we add a region merging process to deal with over-segmentation (watershed is prone to over-segmentation when addressing difficult images). In particular, the proposed region merging process is guided or constrained by the individual cell segmentations of previous reference planes.

The following rules are applied to detect when the splitting process is necessary: 1) if the same cell is marked with a collision followed by a separation, we process all the temporal steps between the marks; 2) if a cell separation is not preceded by a collision mark, we process every temporal step back to the beginning of the volume; and 3) if a collision mark is not followed by a separation one for the same cell, we process every temporal step forward to the end of the volume.

Hence, the output of this first pass is a set of refined segmentations $\mathbf{C}_t = \ t \in [1, T]$ where collisions have been identified and their segmented regions split into the constituent cells.

3.2.2. Pass 2. Second Kalman pass and morphological post-processing

The second step starts with the use of the same 3D Kalman filter over the refined segmentations which are now free from collisions. Then, a morphological post-processing aims to remove any short-term temporal data instability (resulting in spurious segmentations) by taking into account the rest of cell instances in the trajectory. In this manner, the morphological post-processing corrects the segmentation of those cells detected as outliers in a particular trajectory. Figure 6 illustrates this process.

3.2.3. Pass 3. Third Kalman pass and final trajectories

The last step of the process entails applying the same 3D Kalman filtering process to the final segmentations, attained after handing collisions and removing short-term temporal outliers. As a result of this third step, we obtain the final trajectories.

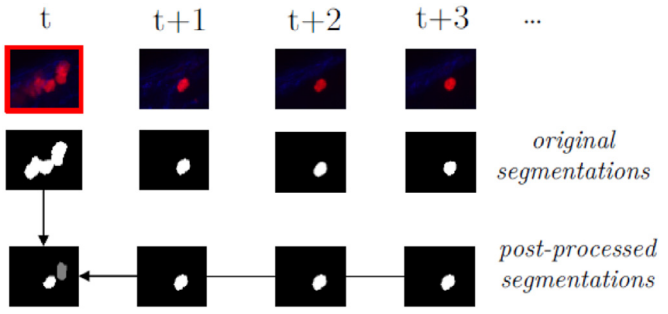


Fig. 6. 3D tracking. Illustrative example of how the morphological post-processing works when outliers are detected. In red, time instant exhibiting a detected outlier. From top to bottom, original segmentations and post-processed ones, adapted to the typical volume of the cells constituting the trajectory. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The resulting 3D tracking method is remarkably consistent because it relies on temporal stability to refine the segmentations (collision detection and handling, and short-time temporal outlier removal processes). Moreover, it is robust due to the Kalman filtering, which does not make strong assumptions regarding cell size, shape or trajectory direction.

3.3. Feature extraction

Few studies so far, as the by Henry et al. (2013), have previously tried to characterize neutrophils from automatic measurements. However, their method performs a directional analysis on the basis of a manually-defined lesion area and it is based on few automatic measurements (neutrophil velocity, direction and angle towards the lesion area). In this paper we propose to perform an unrestricted analysis that relies on the specific geometry defined by the blood vessel, and extracts many features accordingly. It is unrestricted because we do not require any manual input and, furthermore, we leverage the geometry of the blood vessel to conveniently align the feature extraction process.

We model the migration process within the blood vessel in two different coordinate systems: a) Polar coordinates: as the 3D blood vessel segmentation produces a temporal sequence of quasi-elliptical cross-sections, each cell position can be referred to this cross-section in polar coordinates (angular and radial position), i.e., the cell is located within the elliptical cross-section in terms that naturally fit the shape of the vessel; and b) Vessel-aligned Cartesian Coordinates: in this case, we first set the Cartesian coordinate system so that the x-coordinate is aligned with the bloodstream direction, and y-coordinate corresponds to the major axes of the elliptical cross-sections. Then, the cell migration orientation is encoded in differential terms with respect to the bloodstream direction $\Delta\boldsymbol{\theta}^n$ (see Fig. 7). This is particularly appropriate for cell migration, because, apart from shape changes, most of the biological mechanisms result in either radial motility or erratic trajectories.

To our knowledge, this is the most exhaustive set of automatic features computed from microscopy images so far. We aim to model short-term behaviors of the cells based on instantaneous and dynamic features. Instantaneous features are extracted independently for each time instant, while dynamic features are computed over the whole trajectory. Some of the instantaneous features are related to the physical attributes of cells, such as size and shape. Others describe the relative position of the cells with respect to the blood vessel, such as their polar position or their minimum distance to the vessel. Additionally, we consider the well-

Table 1
Proposed set of features for cell migration description.

Type	Attribute	#	Feature name
Instantaneous	**Size/shape**	1	Volume
		2	Superficial area
		3–5	Height, maximum width and height/width rate
		6–8	Sphericity, prolate and oblate ellipticity
		9–11	Principal axes length
		12–13	Extent and solidity
		14	Equivalent diameter
	Cell/blood vessel position	15–17	Cell X/Y/Z axis orientation respect to the bloodstream
		18–19	Cell polar position (radius/angle)
		20	Distance from cell center to blood vessel surface
		21	Minimum distance from cell to blood vessel surface
	Reliability	22	Well-segmented cell probability (pth)
Dynamic	**Motility**	24	Trajectory length
		25	Traveled distance
		26–27	Maximum and mean velocity
		28–29	Meander ratio and tortuosity
		30	Angle between the trajectory and bloodstream
	Size/shape	31–44	Mean of size/shape instantaneous features (1–14)
		44–57	Standard deviation of size/shape instantaneous features (1–14)
	Cell/blood vessel position	58–62	Mean of cell position instantaneous features (15–19)
		63–67	Standard deviation of cell position instantaneous features (15–19)
		68–69	Mean of cell/blood vessel position instantaneous features (20–21)
		70	Minimum distance from the trajectory to blood vessel surface
	Reliability	71–74	Mean, maximum, minimum and range of well-segmented cell probability (22)

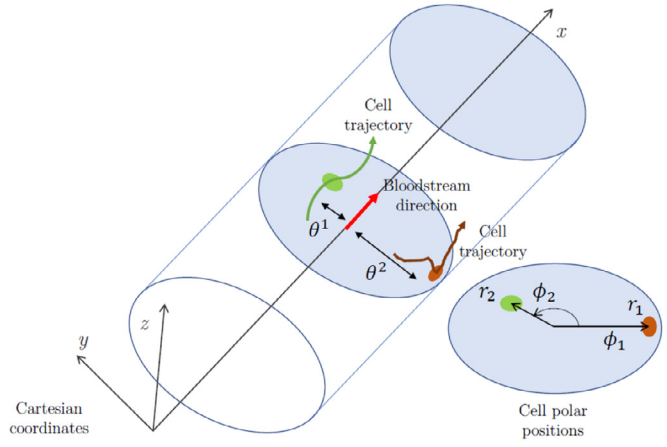


Fig. 7. Representation of cell migration. For each cell, we extract its position in polar coordinates (radius r and angle ϕ) with respect to the quasi-elliptical cross section of the blood vessel and, additionally, describe its trajectory orientation in differential terms with respect to that of the bloodstream: $\Delta\boldsymbol{\theta}^n = [\Delta\theta_x^n, \Delta\theta_y^n, \Delta\theta_z^n]$.

segmented probability estimation p_t^n (see Section 3.1) to include certain information regarding the reliability of the description in the feature set.

The computation of dynamical features requires considering a minimum-length trajectory segment to be considered meaningful and stable. In our experiments we have selected $\Delta_t = 21$ time instants as the minimum-length of a short-term trajectory, following our experimental protocol (Sreeramkumar et al., 2014). In particular, we use a Hann window centered on the time instant to be analyzed, and a temporal spanning of Δ_t instants. Although it is possible that a cell changes its behavior over this short-term time interval (for example, it might start migrating at any instant), we assume that these changes do not happen within a segment. Instead, we consider that other segments of the same cell may show different behaviors.

Table 1 summarizes the complete set of features proposed in this paper for short-term cell characterization. In summary, we propose a set of 74 features, which can be either instantaneous or dynamic, mainly based on essential attributes of the cells related to their size, shape, motility and relative position with respect to the blood vessel.

3.4. Cell selection module

Although the proposed 3D strongly-regularized segmentation module is quite robust, some of the cell segmentations may not be precise enough, with a negative impact on the subsequent cell behavior analysis. In particular, some segmentations could not be suitable to extract features from, and thus, they could provide noisy features for the subsequent analyses. Hence, ACME also includes a cell detection module, to detect those cells and remove them before they influence the explainability.

To properly analyze this potential source of error, we have calculated the correlation between the instantaneous features derived from ground-truth segmentation masks and those derived from our 3D CNN segmentations, for several levels of segmentation quality (measured in terms of IoU, *Intersection over Union*, between proposed segments and ground truth annotations). Figure 8 represents these correlations for all the proposed instantaneous features (#1 to #21 in Table 1) and for several levels of segmentation quality. As can be inferred from the results, only when IoU is 0.5 or higher, the features extracted from the CNN-segmented cells show a sufficient correlation with those coming from ground-truth segmentations. A direct consequence from this study is that our system requires an additional module in charge of automatically identifying which of the tracked cells correspond with well-segmented ones (and therefore will produce an IoU over 0.5), in the absence of ground truth annotations.

This cell selection module relies on the complete set of instantaneous and dynamic features, including that of the well-segmented cell probability, and some expert knowledge from biologists to make its selections. Specifically, it can be broken down into the following sequential stages (see Fig. 9):

1. *Trajectory-length filtering*: in order to guarantee certain stability of the dynamic analysis, we have removed those cells that do not remain at least $\Delta_t = 21$ time instants, in accordance with (Sreeramkumar et al., 2014).

Fig. 8. Correlation between instantaneous features derived from ground-truth segmentation masks and those derived from 3D CNN segmentations, as a function of IoU. As can be observed, for most of the features, correlation is strong (higher than 0.5) when $TH_{iou} \geq 0.5$. It should be also noticed, however, that those features related to the cell orientation with respect to that of the bloodstream (15–17) exhibit a higher sensitivity to TH_{iou}.

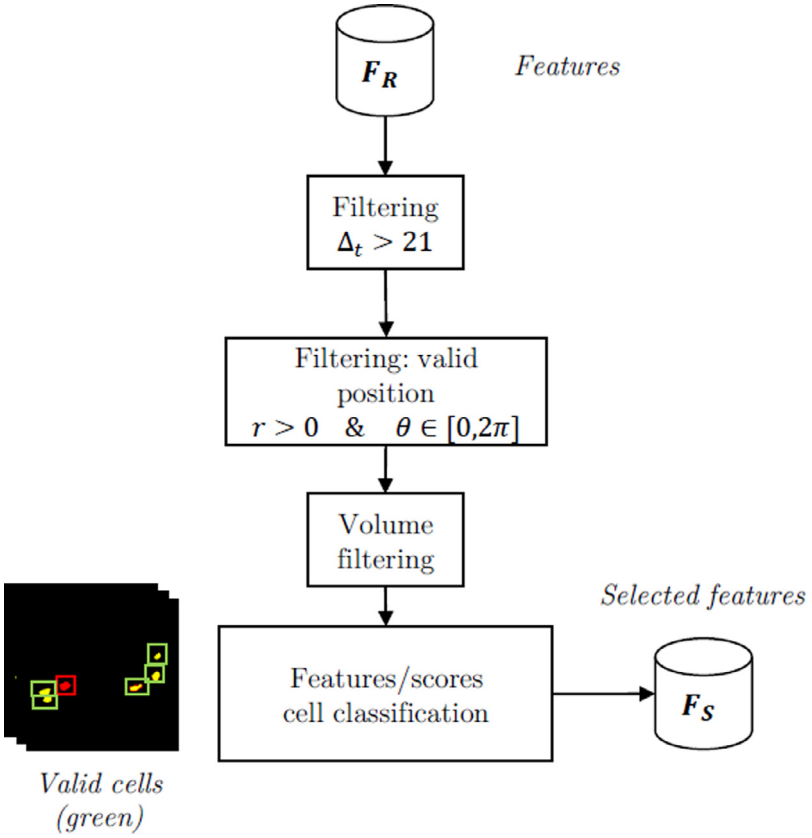


Fig. 9. Flowchart of the cell selection module. After the feature extraction, we apply trajectory-length, low-confidence position and volume filters to remove the regions not susceptible to be cells. Finally, we apply a classifier to select the well-segmented cells with high precision.

2. *Low-confidence position filtering*: some cells, especially when they are at the ends of the volume, cannot be located with accuracy due to incomplete blood vessel cross-sections, thus, resulting in low-confidence position estimations. These cells are also removed from further studies.
3. *Volume filtering*: even after collision detection and handling modules, there are still a few agglomerations or badly segmented cells left. Therefore, those segmented regions whose size remains notably different from the typical cell size,sz, considered by biologists ($< 0.5sz$ and $> 1.75sz$) are also removed from further analysis.
4. *Well-segmented cell classifier*: finally, features of the surviving cells are fed into a Random Subspace classifier (Ho, 1998) that is trained to select those cells with enough segmentation quality, i.e., IoU higher than 0.5. The classifier has been trained by labeling a subset of the data and the decision threshold has been selected to ensure that the precision of cell-detection is above 95% (i.e., at least 95% of the cells in the feature database are properly segmented, with a IoU greater than 0.5).

4. Hierarchical explainability

The automatic feature extraction process proposed in this paper aims to become a powerful tool for biologists to characterize cell migration and, therefore, to gain insight into those biological processes where migration plays a relevant role (García-Prieto et al., 2017; Barabino et al., 2010). In this section, we describe a system that demonstrates the usefulness of ACME. In particular, we assume that the 4D microscopy data come from different populations or groups which are known (each one representing a group of interest, e.g., a control group, a mutation, a therapy, etc.), and the cell behaviors can be described as latent distributions over these groups, i.e., each group contains several cell behaviors, each one in a different proportion. The assumption of having latent behaviors over observable populations is applicable to multiple biological scenarios without loss of generality (Ballesteros et al., 2020; Di Pilato et al., 2021b). Hence, for each cell under analysis, we know the group it belongs to, but ignore its behavior. Moreover, the same cell may exhibit different behaviors over time, which makes our analysis to focus on short-time periods where the cell behavior is stable (see the definition of Δ_t in Section 3.3).

Once we have extracted the features representing each temporal segment of a cell, we aim to group them and discover consistent behaviors among cells, and to explain the resulting behaviors based on their most discriminative features. To this purpose, we have designed a hierarchical explainability module, which operates as follows (see left part of the Figure 10): first, it groups the cells into consistent behaviors using a clustering algorithm over the feature vectors introduced in Section 3.3; second, it organizes these behaviors in a hierarchy (from easily distinguishable behaviors –or groups of behaviors– to barely distinguishable ones); and finally, it explains the resulting behaviors in terms of their most discriminative features. In the following paragraphs we describe these procedures in more detail:

1. *Non-supervised behavior discovery*. The behavior discovery module is in charge of selecting the optimal number of behaviors K_{opt} present in the data and clustering them into K_{opt} clusters. The number of clusters K is the main parameter of our clustering algorithm, K-Means (Lloyd, 1982), and has been chosen as follows: we run the algorithm for different values of $K \in [2, K_{\max}]$, i.e, for different number of potential clusters/behaviors, and a histogram of the number of cells exhibiting each behavior is computed for each group. Then, the optimal number of behaviors is the one which minimizes the histogram intersection between the two most similar groups as follows:

$$K_{opt} = \underset{K\in[2,K_{max}]}{\arg\min} \left[\max_{\substack{i,j\in G \\ i\neq j}} HI\left(\mathbf{h}^K_{g_i}, \mathbf{h}^K_{g_j}\right) \right] \tag{9}$$

where G represents the set of groups and $\mathbf{h}^K_{g_i}$ is a K-dimensional vector with the normalized histogram of behaviors for the group i:

$$\mathbf{h}^K_{g_i} = \begin{bmatrix} h^1_{g_i} & h^2_{g_i} & \dots & h^K_{g_i} \end{bmatrix} \tag{10}$$

with h^k_i being the proportion of cells with behavior k in the group g_i.

By minimizing this histogram intersection, we separate behavior distributions in the considered groups. For example, if the goal is to distinguish between a healthy and a pathological group, the selected K_{opt} must be chosen to maximally separate these groups in terms of their distribution of behaviors. After K_{opt} has been selected according to this criterion, a simple but high-dimensional effective K-Means algorithm (Lloyd, 1982) is used to cluster the normalized features into K_{opt} behaviors.

2. *Hierarchical explainability*. We have designed a hierarchical approach to achieve explainable behaviors mainly due to the fact that the degree of dissimilarity is often variable between different behaviors. Our aim is to hierarchically arrange the behavior space, starting with the two most different subsets of behaviors (i.e., those subsets of behaviors which lay more separated in the feature space) and, then, performing subsequent divisions according to subtler differences between behaviors. To do so, a modified Binary Partition Tree (Breiman et al., 1984) has been chosen to hierarchically arrange the behavior space. At each branch of the tree, a binary partition is chosen as follows:
   a) Considering only the subspace of behaviors contained in its parent branch, we build all the combinations of potential binary partitions of that set. Each candidate partition entails two subsets of behaviors s_1 and s_2 whose division is evaluated.
   b) For both subsets $s_i, i = 1, 2$, and assuming independence among features, we model the distribution of each feature f as Gaussian Variables $p_{s_i}(f) = \mathcal{N}(f|\mu_{f,s_i}, \sigma_{f,s_i})$.
   c) We compute the overlapping between subsets as the median over the features of the overlapping coefficient (Inman and Bradley, 1989) between the distributions of subsets s_1 and s_2:
   $$OVL(s_1, s_2) = \underset{f}{\text{median}}\, ovc(p_{s_1}(f), p_{s_2}(f)) \tag{11}$$
   where *ovc* stands for overlapping coefficient between Gaussian distributions.
   d) Among the candidate partitions, we choose that one s^*_1, s^*_2 that minimizes OVL:
   $$s^*_1, s^*_2 = \underset{s_1,s_2}{\arg\min}\, OVL(s_1, s_2) \tag{12}$$
   e) Finally, the algorithm determines if this best subset of behaviors is better than each behavior considered separately, in terms of Davies-Bouldin criterion (Davies and Bouldin, 1979). If so, the hierarchy is updated with this partition and the algorithm continues evaluating new partitions. Otherwise, the algorithm ends with a branch in which the original subset is finally partitioned into individual behaviors. Hence, it may happen that the last step of the hierarchy is not a pure binary but a multiple partition into the remaining behaviors.

Once the hierarchy **H** has been obtained, we use a L1-regularized Logistic Regressor to select the most important features at each decision level. It is known that L1-regularization

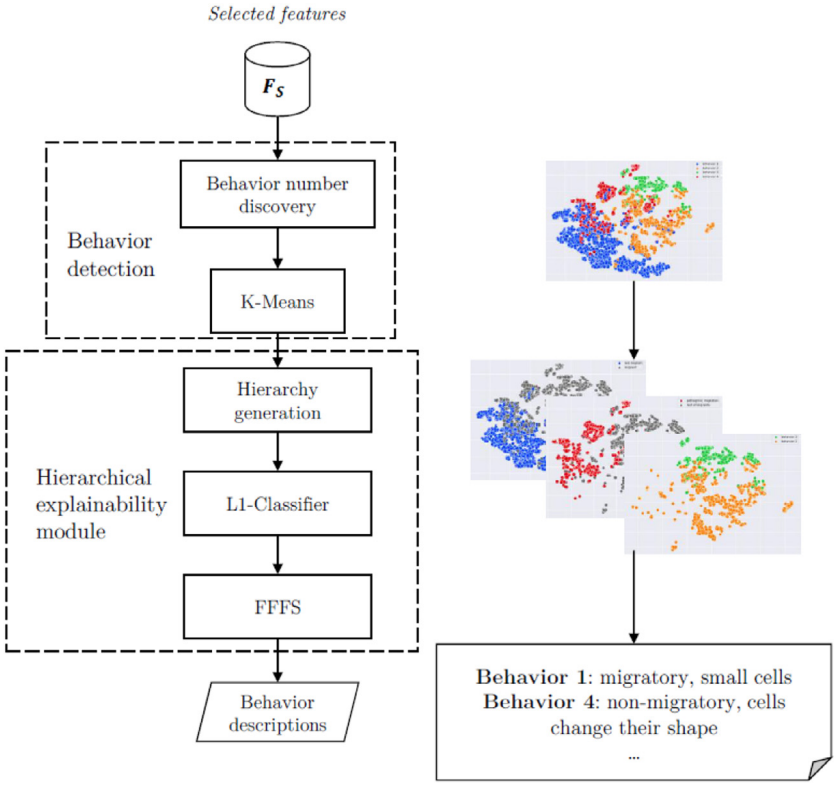


Fig. 10. Block diagram of the hierarchical explainability module. The goal of this module is to partition the data into a small number of behaviors, which are illustrated in different colors on the right part of the figure. This module operates sequentially (see left part of the figure): 1) it groups the cells according to their behaviors; 2) it arranges the behavior space in a hierarchy; and 3) it explains the resulting behaviors by identifying the most relevant features.

(Lasso) is useful for feature selection as it tends to set weights of irrelevant features to zero. In particular, we have chosen the weight of the Lagrangian regularization parameter as the minimum one that obtains a relative faithfulness in the training data over the 98% with respect to a non-regularized classifier (i.e., the classifier accuracy at the corresponding level is at least 98% of the overall accuracy when no regularization is used, and therefore the classifier optimally fits the training data). In consequence, for each decision node, we have only a few features weighted by non-zero coefficients. Furthermore, to avoid any kind of redundancy (which might detract from their interpretability), we apply an additional Feed-Forward Feature Selection (FFFS) stage.

At the end of this process, we can characterize each behavior with a small number of features together with an indicator of their importance, which is derived from the corresponding weight of the L1 classifier.

5. Experimental results

5.1. Dataset description and experimental setup

To assess ACME, we have built an intravital microscopy (IVM) imaging database containing neutrophils migrating within venules in cremaster muscles of mice, containing 147 4D captures composed of 2334 3D volumes. Data were acquired from 23 mice, a similar number to the one in the previous studies (Sreeramkumar et al., 2014; García-Prieto et al., 2017), with an IVM system built by 3i (Intelligent Imaging Innovations, Denver, CO, USA) on an Axio Examiner Z.1 workstation (Zeiss, Oberkochen, Germany) mounted on a 3-Dimensional Motorized Stage (Sutter Instrument, Novato, CA, USA). This set up allows precise computer-controlled lateral movement between XY positions and a Z focusing drive for confocal acquisition. 4D captures were acquired at 0.667×0.667 micrometers X/Y-intervals with a spatial resolution of 256×256 pixels, and variable Z-intervals ranging from 1 to 2 μm. The temporal sampling period was also variable, ranging between 4 and 20 s. Hence, prior to any study, all the extracted features were normalized in terms of both Z and temporal sampling period.

The 4D captures come from four populations or groups, namely:

- **Group 1, wild-type/control**: mice without a treatment. Biological hypotheses suggest that neutrophils belonging to this group are larger, change their shapes and migrate over blood vessel surface.
- **Group 2, anti-Plt**: mice with platelet depletion. Our existing data suggested that neutrophils belonging to this group are smaller, spherical and, hence, they do not migrate.
- **Group 3, FGR-KO**: mice transplanted with bone marrow of a knockout mutant for FGR protein. The neutrophils in this group are expected to behave intermediate in relation to those of the groups 1 and 2.
- **Group 4, FGR-INH**: mice treated with an inhibitor of the FGR protein. Neutrophils in this group should behave in a similar way to those of group 3.

The dataset has been split into training, validation and test sets. Moreover, 120 3D volumes belonging to groups 1, 2 and 3 have been annotated pixel-wise (blood vessel and 794 neutrophil instances) to train the 3D segmentation CNN and the selection module. Each one of the annotations has been prepared by one expert, but the level of variability among experts is reduced due to the annotation procedure, based on thresholding with a discrete number of valid thresholds and manual refinement. Specifically, the Fleiss' kappa coefficient κ (which measures the inter-annotator agreement) (Fleiss, 1971) is above 0.78 for five different annotators. Taking into account that Fleiss' kappa ranges between $\kappa = -1$ (to-

Table 2
Comparative of blood vessel segmentation results with state-of-the-art methods.

Method	**IoU** (%)
U-Net 3D (Çiçek et al., 2016)	81.35
Deeplab V3-3D (Chen et al., 2017; Choi, 2020)	80.86
ACME-vessel-only (cell-aware)	**88.85**
ACME	**88.09**

tal disagreement between the ratings) and $\kappa = 1$ (total agreement between the ratings), we conclude that these levels of variability between annotations are not relevant.

In addition, in the annotated set, there are 149 temporal instances with collisions or separations out of a total number of 794 temporal instances of cells, and with very varied dynamics. Therefore, we consider the size of the dataset enough to address the challenge of collision management.

Groups have been divided into 4 folds to perform cross-validation, and both the segmentation CNN and the selection module (the trainable modules) have been trained for each fold in an individual way. For feature extraction, the remaining data from groups 1, 2 and 3 have been tested on their corresponding test fold, while data from group 4 have been tested using a system trained on the complete set of annotated data.

The code has been developed using Python (for 3D CNN segmentation and data visualization) and Matlab (Mathworks^{TM}, USA) for the rest of modules.

The experiments in this section are organized as follows: first, we present the results of the proposed system in terms of segmentation and tracking performance, which are compared with those of several state-of-the-art systems. For every one of them, we perform specific ablation studies, showing how the proposed contributions gradually increase the performance of the methods. Apart from that, we present an incremental ablation study to analyze how the precision and recall in detection vary along the pipeline. Second, we examine the explainability results and their consonance with biological hypotheses. Finally, we perform an error analysis and discussion, and a study of the influence of the operation point of the system.

5.2. Comparison with the state-of-the-art

The goal of this section is to experimentally assess the proposed system in comparison with other relevant approaches of the state-of-the-art. We have organized the results into two subsections: first, the performance of 3D strongly-regularized segmentation CNN is assessed; then, the performance of 3D tracking is evaluated. Additionally, complementary results in terms of precision and recall are given throughout the complete pipeline.

5.2.1. 3D strongly-regularized segmentation module

The first set of experiments is devoted to assess the 3D strongly-regularized segmentation network in comparison with other state-of-the-art methods for medical image segmentation. Semantic segmentation results are shown in terms of IoU in Tables 2 and 3 for the blood vessel and the cells, respectively. Table 3 also shows the results in terms of precision and recall for the cell detection task.

Table 2 shows the results for blood vessel segmentation. For comparison with the state-of-the art we have selected U-Net 3D (Çiçek et al., 2016) and Deeplab V3-3D (Chen et al., 2017), (Choi, 2020), state-of-the-art approaches for conventional semantic segmentation. As it can be observed, ACME provides a 10% improvement in terms of IoU, allowing us to conclude that cell positions relative to the vessel are very important for the segmentation process. Furthermore, when comparing the two versions of our system, there no significant difference between ACME-vessel-only segmentation and ACME using the joint approach (blood vessel and cell segmentation). Nevertheless, it should be noticed that for our vessel-only segmentation system to work properly, we needed to regularize with the cell labels (the cells within the blood vessel were annotated). Therefore, although the joint segmentation was not revealed to be essential for the vessel segmentation (as it will be for the cell segmentation), the vessel-only segmentation required cell-related information to be successful.

Table 3 shows the results for cell segmentation and detection. Although we have designed our 3D CNN as a semantic segmentation system (e.g. it does not discriminate between joint instances of cells), mainly because our robust 3D tracking method can handle collisions, some state-of-the-art approaches are formulated as instance segmentation networks. For comparison purposes, we have selected some specific approaches for cell detection: Retina U-Net (Jaeger et al., 2018), QCA-Net (Tokuoka et al., 2020) (a very recent instance segmentation approach) and the original version of U-Net (as a semantic segmentation approach). Apart from these state-of-the-art approaches, we have included three versions of our 3D system to reveal the aggregated impact of each of our contributions on top of the U-Net 3D backbone, namely: the focal loss (ACME-Focal), the auxiliary segmenter (ACME-Focal & Aux), and the well-segmented cell probability estimation (ACME).

ACME performs much better than both Retina U-Net (even with anchors adapted to our particular problem) and QCA-Net. It is worth mentioning that QCA-Net is very effective in terms of precision due to its detection network, but it produces slightly worse segmentations in demanding scenarios as the one posed in this paper (low recall in our task). Regarding the three versions of our system, the focal loss (ACME-Focal) makes the network to pay attention to the cells, substantially improving recall with respect state-of-the-art approaches; the proposed auxiliary segmenter (ACME-Focal & Aux) improves precision in cell detection in a 10% while maintaining the recall; and finally, the estimation of the probability of well-segmented cell, which completes the system, only produces a slight improvement in precision and recall. In any case, the purpose of this last subsystem is to serve the subsequent cell selection module, and so its impact will be significant to improve the precision even more.

Finally, we provide some illustrative examples of both blood vessel and cell segmentation in Figs. 11 and 12, respectively. Results for blood vessel segmentation show that U-Net 3D poorly generalizes and when the blood vessel contrast is low cannot segment the venule. However, ACME as well as ACME-vessel-only cell-aware, which is not included in the figure, produce consistent and precise results. Regarding the cell segmentation results, Fig. 12 shows that ACME-Focal is less precise in cell segmentation due to three reasons: 1) ACME benefits from blood vessel segmentation to discard cells out of the blood vessel; 2) ACME produces more compact segmentations and less spurious detections; and 3) ACME is more robust when contrast luminance is not stable along the blood vessel.

5.2.2. Three-pass 3D cell-tracking with collision handling

In order to assess the proposed 3D tracking system with collision handling, we compare its results with those of PhagoSight (Henry et al., 2013) in terms of:

- Recall: number of cells correctly associated with their ground-truth trajectory with respect to the total number of ground-truth cells.
- Precision: number of cells correctly associated with their ground-truth trajectory with respect to the total number of detected cells.

Table 3
Comparative of cell segmentation and detection results with state-of-the-art methods. *Retina U-Net was adapted to our task: anchor sizes and feature map depths where the anchors are extracted were modified.

Method	**IoU** (%)	Precision (%)	Recall (%)
QCA-Net w/o NDN (Tokuoka et al., 2020)	45.85	28.65	25.31
QCA-Net (Tokuoka et al., 2020)	45.53	60.82	22.29
Retina U-Net* (Jaeger et al., 2018)	51.55	46.46	56.56
ACME-Focal	62.09	56.32	75.19
ACME-Focal & Aux	62.82	65.81	74.81
ACME	**64.44**	**67.13**	**78.46**

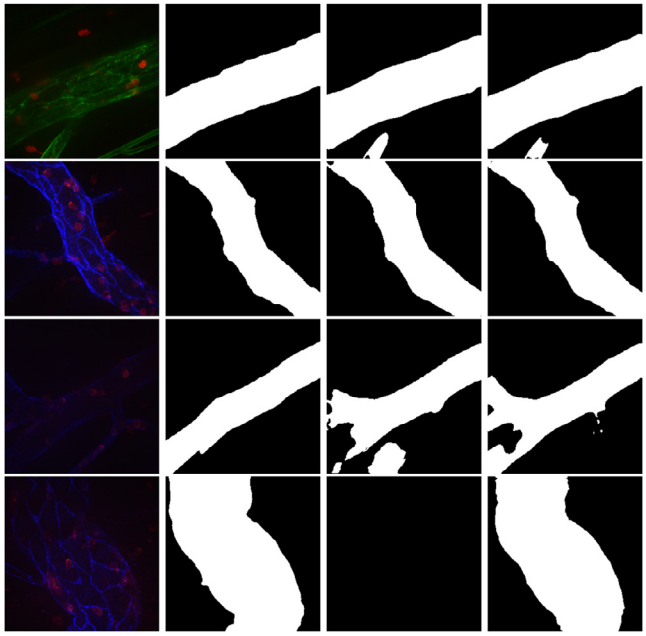


Fig. 11. Illustrative examples for blood vessel segmentation (z-stacked). First column: original volume; second column: ground-truth venule segmentation; third column: U-Net 3D and fourth column: ACME.

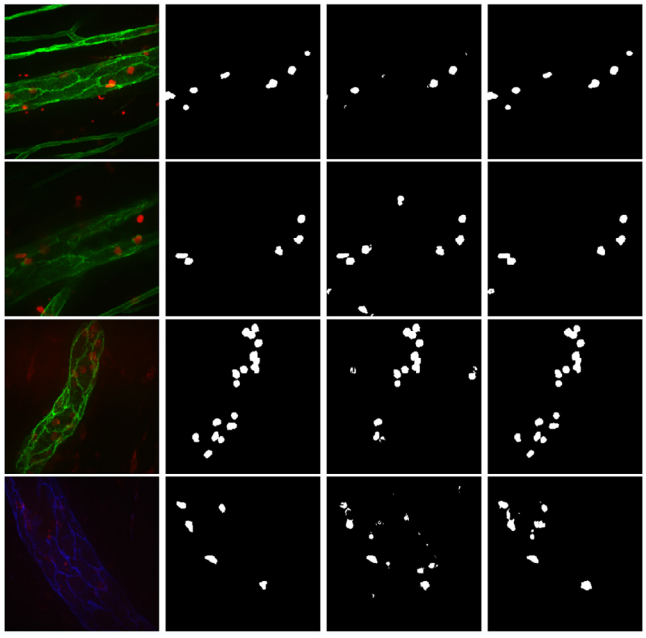


Fig. 12. Illustrative examples for cell segmentation (z-stacked). First row: original volume; second row: ground-truth cell segmentation; third row: ACME-Focal and fourth row: ACME.

- IDE: track identification instantaneous errors, percentage of cell trajectory errors for each detected track.
- CMR: collision management success ratio, i.e., percentage of cell collisions correctly splitted.

Results are shown in Table 4. As observed, IDE is notably lower for our tracking system, which indicates that the tracks obtained by ACME are more stable than in the compared approaches. Furthermore, CMR increases, thus, more collisions have been correctly separated. Finally, when comparing precision and recall of the segmentation system before and after the 3D tracking system and collision management, there are no significant differences. The decay in precision and recall of Phagosight and ACME is due to over-segmentation of some individual cells (much less frequent in ACME). Finally, Fig. 13 shows several examples of collision handling which illustrate the better performance of ACME with respect to PhagoSight (Henry et al., 2013).

5.3. Support to biological hypotheses

ACME obtains substantially better results for both 3D segmentation and tracking than the considered state-of-the-art methods, as demonstrated in the previous section. However, to perform a robust behavioral analysis precision should be kept very high. Neutrophil selection module, as described in Section 3.4, removes invalid cells for subsequent analyses. Improved results in terms precision and recall after the filtering processes and Random Subspace classifier are shown in Table 5.

Neutrophil selection module allows the system to reach the 95%-precision operation point, which, as previously mentioned, allows us to perform a robust short-term behavior analysis of the different populations, not biased by wrongly detected non-cell regions or artifacts. As shown in Table 5, this high performance level in terms of precision is achieved in exchange for losing recall, which is less relevant from the point of view of our analysis as long as we can ensure the availability of enough cells to perform that analysis (see data below). It is worth noticing that the most significant drop in recall comes from removing those cells whose trajectory is not long enough, a decision based on biological recommendations to guarantee a correct dynamical analysis. Finally, the last significant improvement in terms of precision comes from the Random Subspace classifier, which also causes a decline in recall.

As a result of the application of the selection module, ACME has extracted features from 6800 time instances belonging to more than 400 neutrophils with very high precision. Data distribution is shown in Table 6. At this point, the hierarchical explainability method (described in Section 4) has been applied to these data in order to analyze its agreement with our experimental-based hypotheses. Let us briefly recall that this module comprises two stages: 1) behavior detection: discovery of the optimal number of cell behaviors to represent the data; and 2) hierarchical explainability: creation of a behavior partition tree on which the explain-

Table 4
Cell tracking results with respect to PhagoSight (Henry et al., 2013).

Method	Precision (%)	Recall (%)	**IDE** (%)	**CMR** (%)
Segmentation data	**67.13**	**78.46**	-	0
PhagoSight (Henry et al., 2013)	61.36	70.25	3.36	12.15
ACME	**66.45**	**75.67**	**1.58**	**31.40**

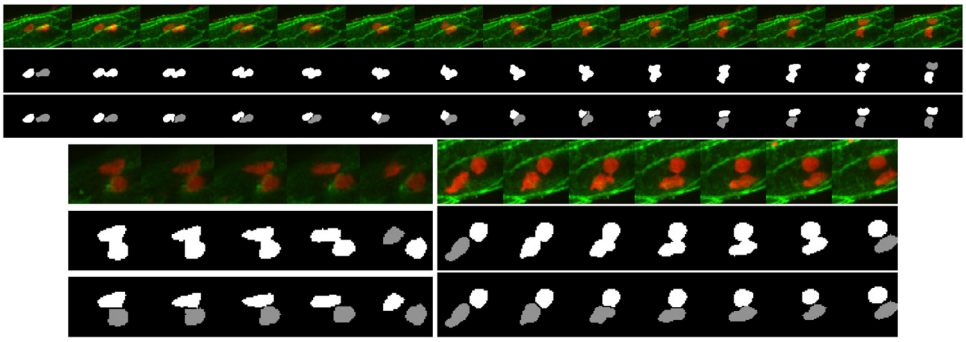


Fig. 13. Illustrative examples of collision detection in tracking (temporal sequences). First row: original volume; second row: PhagoSight (Henry et al., 2013) results; and third row: ACME results. As can be seen, ACME is able to split temporarily touching cells by relying on reference segmentations and does not swap trajectories between cells.

Table 5
Cell detection performance of the proposed system in terms of precision and recall. The impact of every component of the neutrophil selection module is evaluated taking as reference the baseline results achieved by the previous modules (segmentation and tracking).

Component	Precision (%)	Recall (%)
Segmentation data	67.13	**78.46**
Tracking data	66.45	75.67
Trajectory-length filtering	85.12	38.92
Valid-position filtering	85.56	38.79
Volume filtering	85.56	38.79
Random Subspace classifier	**95.28**	**30.48**

Table 6
Dataset description.

Group	# Captures	# Neutrophils	# Time instances
1	48	169	2685
2	32	138	2008
3	16	40	615
4	27	95	1492
Total	54	442	6800

ability will be based on, and selection of the most relevant features explaining each decision level of the tree.

The behavior detection module found five behaviors in these data. The criterion chosen to determine the optimal number of behaviors, K_{opt}, was to maximize the difference between the behavior distributions of groups 1 (control) and 2 (anti-Plt), which are, *a priori*, the two most opposed populations. Figure 14 (left) displays a t-SNE (van der Maaten and Hinton, 2008) graph that arranges the resulting behaviors in a 2D space (t-SNE allows the representation of high-dimensional data, 74-D in our case, in a low-dimensional space). Figure 14 (right) shows a stacked histogram illustrating the behavior distribution within each group. These stacked histograms reveal that every behavior is present in all groups, but in different proportions. For example, the first group contains the largest proportion of neutrophils exhibiting the second behavior, while the first behavior is dominant for the second group.

Once we have discovered the behaviors, we have automatically created a hierarchical behavior tree by partitioning the data at each level into the two most distinguishable behaviors (or groups of behaviors). Figure 15 goes a step ahead and shows the resulting hierarchical tree and the associated explainability. In particular, for each node, the most relevant features explaining the underlying partition are superimposed on the tree to illustrate the concept of hierarchical explainability. As we delve into the hierarchy, features are less informative and describe lower level differences between behaviors. It should be also mentioned that the features are listed in order according to their relevance.

Beyond the representation of Fig. 15, different hierarchy levels can be easily described in textual form on the basis of the selected features as follows:

- **Level 1: non-migratory/migratory**. This partition distinguish migratory behaviors (2, 3, 4 and 5) from the non-migratory one (behavior 1). In particular, neutrophils exhibiting this non-migratory behavior are more spherical and change less their shape; thus, they travel less distance than the neutrophils of migratory behaviors.
- **Level 2: pathogenic/non-pathogenic**. At this partition level focused on migratory behaviors, behavior 2 (pathogenic migration) is separated from behaviors 3, 4 and 5 (non-pathogenic migration). Neutrophils showing pathogenic migration flatten, move over the blood vessel surface and are bigger than the rest of migratory ones.
- **Levels 3 and 4: non-pathogenic migratory behaviors**. These levels allows us to differentiate behaviors 3, 4 and 5 (all of them non-pathogenic migrations) in terms of size, shape and segmentation system scores (well-segmented probabilities).

In addition to this textual behavior description, we can rely on the most prominent features at each hierarchical level to graphically display how discriminative each feature becomes at that level. Figures 16, 17, 18 and 19 show the most important features to distinguish behaviors at different levels. As can be observed, prominent feature values are highly correlated with behavior distribution, specially at higher levels in the hierarchy. In addition to this, features ranked first at each hierarchy level are more descriptive of the behavior (correlation between feature representations and hierarchy level distributions is higher).

The previous results allow us to conclude the hierarchical explainability method provides insights in full agreement with our biological model. The first group (control) contains the larger proportion of pathogenic migratory neutrophils (larger, moving over

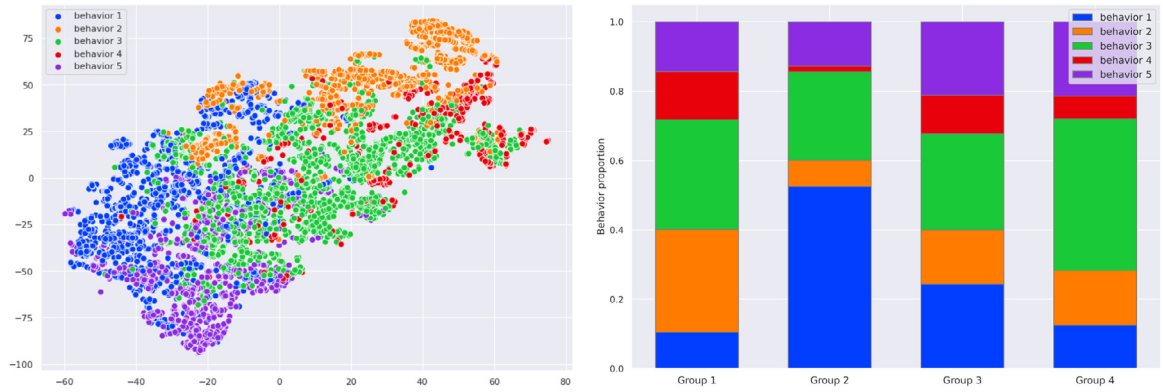


Fig. 14. Spatial arrangement of behaviors in a 2D t-SNE plot (left) and stacked histograms showing behavior proportion within each group (right).

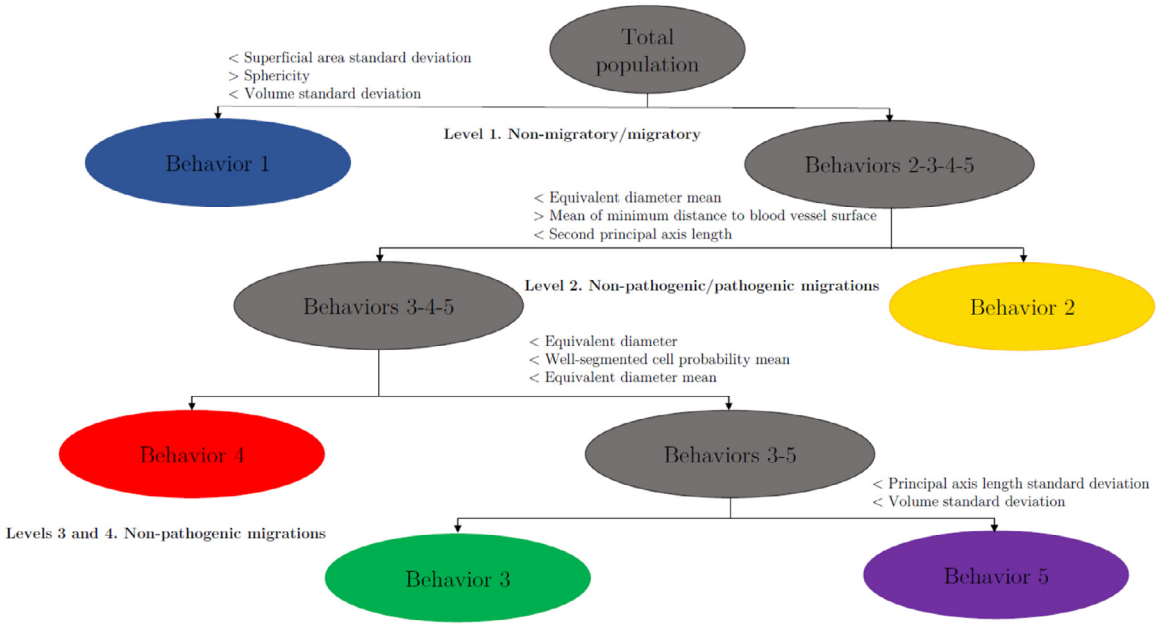
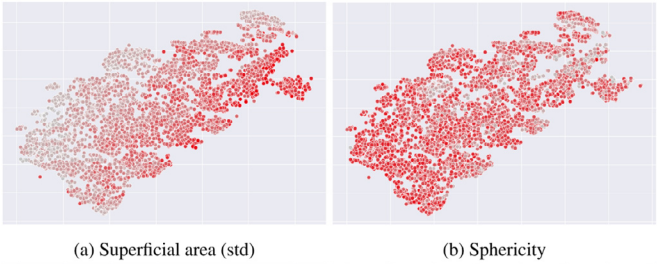


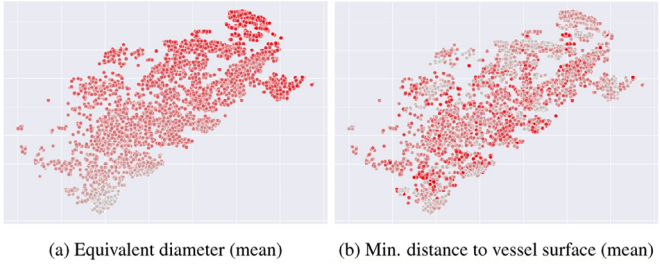
Fig. 15. Hierarchical explainability diagram. Behaviors (or groups of behaviors) are displayed as blobs. For each node, the features selected to explain the partition are shown in order of importance. It should be noted that hierarchical explainability is accumulative, in the sense that, for example, behavior 4 is migratory (level 1), non-pathogenic (level 2) and has the distinctive features explained at level 3.



(a) Superficial area (std) (b) Sphericity

(c) Volume (std) (d) Migratory/non-migratory distribution

Fig. 16. Hierarchical explainability model: t-SNE diagram showing prominent features for discriminating between migratory and non-migratory behaviors.



(a) Equivalent diameter (mean) (b) Min. distance to vessel surface (mean)

(c) Second principal axis length (d) Pathogenic/non-path. distribution

Fig. 17. Hierarchical explainability model: t-SNE diagram showing prominent features for discriminating between pathogenic and non-pathogenic behaviors.

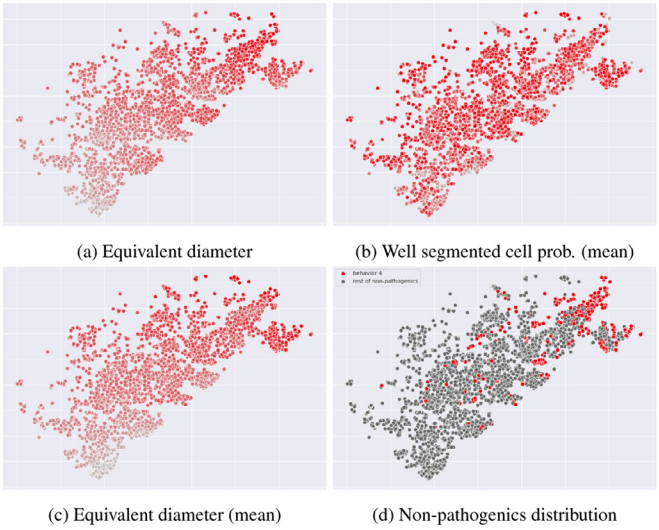


Fig. 18. Hierarchical explainability model: t-SNE diagram showing prominent features for discriminating between behavior 4 from behaviors 3 and 5 (all non-pathogenic).

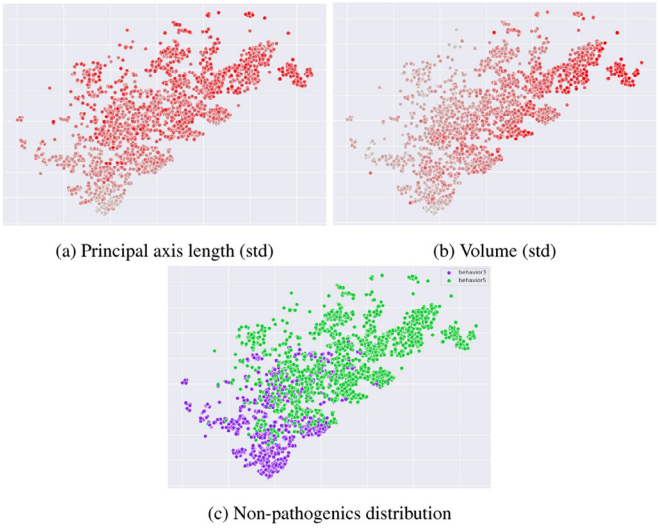


Fig. 19. Hierarchical explainability model: t-SNE diagram showing prominent features for discriminating between behaviors 3 and 5 (both non-pathogenic).

the blood vessel surface and contributing to inflammatory injury), while their abundance becomes notably smaller in groups 3 and 4, revealing the effectiveness of these two therapies. Additionally, group 2 (platelet depletion) contains mainly non-migratory cells: when neutrophils cannot recruit platelets the mechanisms of migration are affected (Sreeramkumar et al., 2014). Non-pathogenic migrations are less frequent in groups 3 and 4 in comparison with the control group, thus, therapies avoid the emergence of the pathogenic migration in favor of the non-pathogenic ones.

5.4. Error analysis and discussion

In order to provide more insight into the capabilities of the proposed system, we have analyzed its errors on a module basis. Moreover, we have examined the sensitivity of the behavior analysis system to the operating point of the cell selection module, in terms of both the stability of the resulting behaviors and the variation of behavior proportions within each group.

Figure 20 shows some illustrative examples of the most serious errors of the blood vessel segmentation branch of the 3D strongly-regularized segmentation CNN. As can be observed, these errors are due to either significant photobleaching affecting the complete volume (or parts of it), or the presence of secondary blood vessels. However, as can be inferred from the examples, even in these cases, the system is robust and usually provides an acceptable result for blood vessel segmentation. Furthermore, the valid position filtering in the cell detection module can remove the cells if they are beyond the limits of a inaccurate segmented blood vessel.

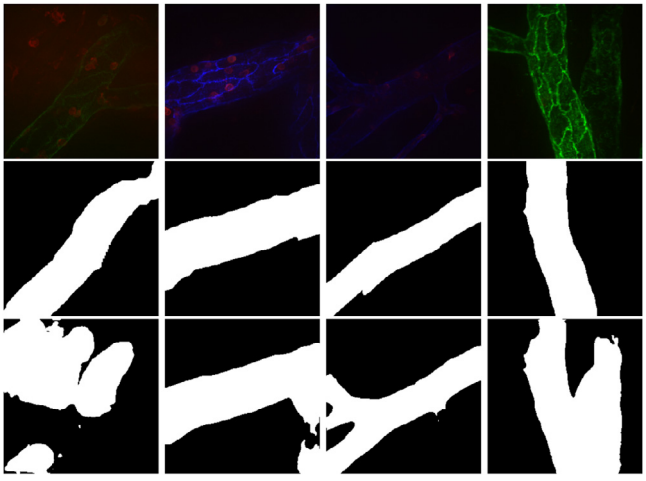


Fig. 20. Illustrative examples of the most serious errors related to the blood vessel segmentation (z-stacked). First row: original volume; second row: ground-truth venule segmentation; third column: ACME result.

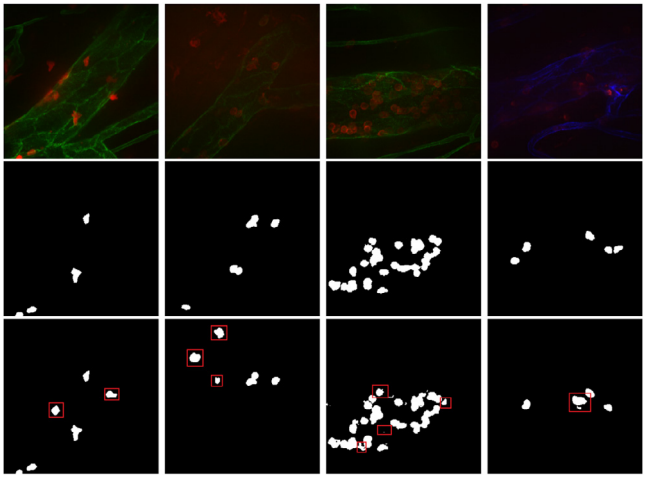


Fig. 21. Illustrative examples of the most serious errors related to the cell segmentation (z-stacked). First row: original volume; second row: ground-truth cell segmentation; third column: ACME result (errors highlighted in red boxes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Figure 21 shows some illustrative examples of the most serious errors of the cell segmentation branch, for example, cells exhibiting poor contrast can result in misdetections. Apart from that, when the blood vessel is not precisely segmented some cells can be segmented outside the venule. Nevertheless, although this is undoubtedly a tough problem, the system is robust enough and produce compact and meaningful segmentations. Moreover, the system is able to segment some multiple-neutrophil agglomerations and even some difficult cells (not annotated in the ground-truth), and the cell selection module selects the best segmented cells, removing some of these errors.

Regarding the 3D tracking system, Fig. 22 illustrates two typical sources of errors where the proposed system fails to split colliding

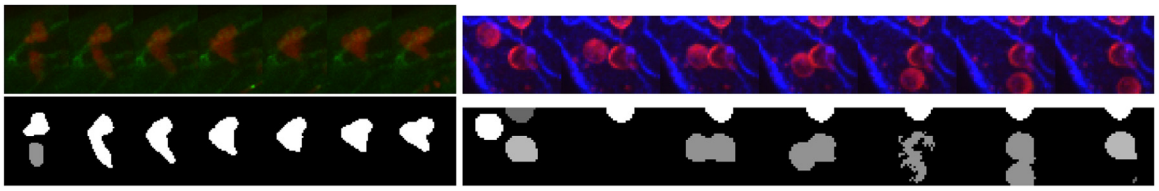


Fig. 22. Two illustrative examples of errors of the proposed tracking system relate to collision management. First row: original volume; second row: ACME results.

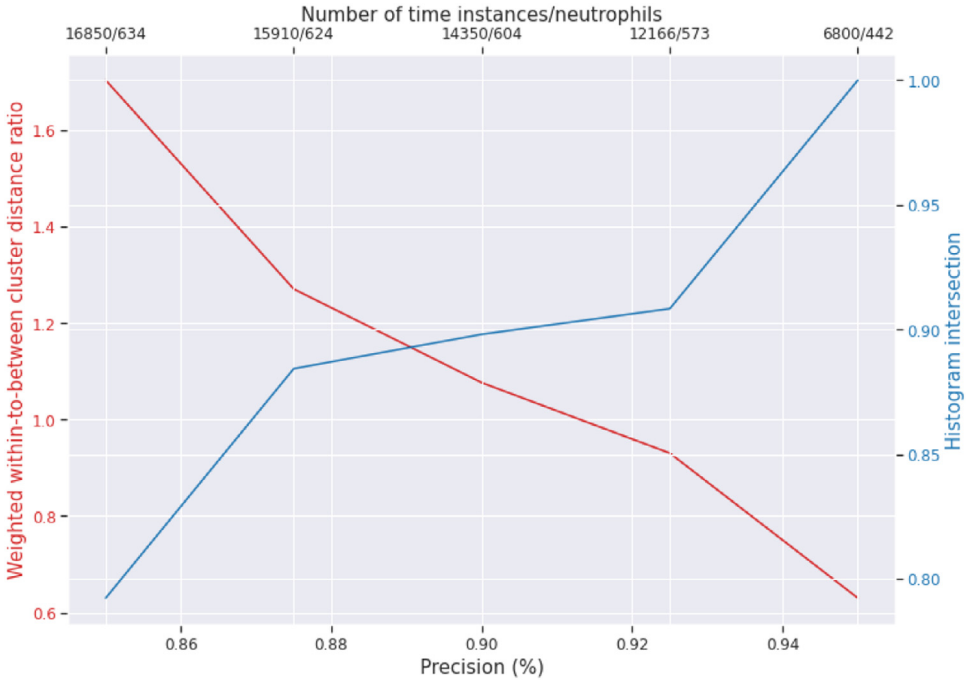


Fig. 23. Sensitivity of the explainability to the precision operation point: quantitative analysis.

cells: 1) the cells collide in a way that the watershed method cannot detect two centers, resulting in only one object; and 2) the 3D strongly-regularized CNN segmentation is not stable across time instances of the cells, so the proposed region merging process cannot be properly constrained by the individual cell segmentations of previous reference planes, and the splitting method struggles to split the touching cells. In any case, these errors do not affect the final results in a meaningful way.

5.5. Sensitivity of explainability to the precision operation point

Finally, we examine the influence of the operation point of the neutrophil selection module on the explainability. In particular, we selected 95% of precision as the system operation point, but this requirement could be relaxed if explainability is not too sensitive to it (this precision threshold could be relaxed to include new samples in the analysis). To study the sensitivity of the system we suggest measuring the variation of two quantities as a function of the precision threshold; namely: the stability of the resulting behaviors and the variation of the behavior proportions within each group.

- *Stability of the resulting behaviors*. Let $\bar{d}_i$ be the intra-cluster dispersion for the ith cluster, and let d_{ij} be the inter-cluster distance between the clusters ith and jth. We define the within-to-between (WTB) cluster distance ratio, D_{ij} as follows:

$$D_{ij} = \frac{\bar{d}_i + \bar{d}_j}{d_{ij}} \tag{13}$$

Let us define a weighted mean distance ratio between clusters as follows:

$$\bar{\mathbf{D}} = \sum_{i=1}^{K_{opt}} \sum_{j=1}^{K_{opt}} \frac{D_{ij} \cdot h^j}{\sum_j h^j} \tag{14}$$

where $\mathbf{h}^j$ is the number of cells for the behavior j.
This weighted mean distance ratio that aims to provide some insight into the separability of the detected behaviors and the stability of the behavior topology in the high-dimensional original space, and we have studied how this measure varies as a function of the precision threshold.

- *Variation of the behavior proportion within a group*. In order to study how the behavior proportion varies as a function of the precision threshold, we propose to compute the histogram intersection within each group between the original case (95% of precision) and the rest of cases (when the precision threshold is lowered).

Figure 23 shows the sensitivity of the system to the precision operation point in terms of the previously described measurements. It should be noted that every decay in precision carry on more instability in behaviors and changes the behavior proportion within the groups, hence the suitability of a 95% precision value. However, above 90% precision, variations in both measurements are manageable. >From a qualitative point view, as can be observed in Fig. 24, the clusters in the t-SNE begin to merge as we lower the precision threshold, specially under 90% precision. Therefore, we can conclude that an operation point higher than

Fig. 24. Sensitivity of the explainability to the precision operation point: qualitative analysis. t-SNE diagrams for several precision operation points. From left to right: 95%, 92.5%, 90%, 87.5% and 85%. It should be noted that for 90% of precision the t-SNE topology begins to change and for 85% the 4th behavior disappears.

90–92.5% precision produces a consistent explainability while the number of neutrophils analyzed notably increases.

6. Conclusion

In this paper, we have introduced ACME, an Automatic feature extraction method for Cell Migration Examination in microscopy imaging that is composed of four modules: 1) a 3D CNN that performs a strongly-regularized segmentation of the blood vessel and the cells; 2) a three-pass 3D cell tracking system with collision handling; 3) a novel and exhaustive feature extraction module which characterizes the short-term dynamics of the cells using coordinate systems tailored to the geometry of the blood vessel; and 4) a cell selection module that guarantees the quality of further analysis based on the extracted features.

The main goal of the system is to fill the gap between imaging-based research and usability in cell migration analysis within blood vessels. Traditionally, 3D cell segmentation has been performed using software such as Imaris or Fiji, where manual supervision is required; thus, the feature extraction process is onerous, and inconsistencies among different annotators are unavoidable. ACME aims to overcome these limitations and provides a single and robust pipeline to segment, track and extract features for this specific task. A direct consequence is a significantly larger amount of available data for the subsequent analysis (up to an order of magnitude larger), in comparison to previous approaches that required human manual participation. ACME is also versatile enough to be applied to different scenarios of 4D cell imaging which follow the described data acquisition protocol, or can be re-trained to obtain the optimal results with other tasks. Furthermore, ACME allows to easily design new features and modify the segmentation architecture to adapt to other scenarios.

We have independently assessed the four modules of the proposed system in a challenging *in vivo* scenario. Our experimental results prove that a joint approach to the segmentation of the blood vessel and the cells leverage the context to notably improve the segmentation performance in comparison with other state-of-the-art methods. In particular, our 3D strongly-regularized cell segmentation approach, improved by the use of a focal loss, an auxiliary segmenter and a well-segmented cell probability estimation, outperforms more complex instance segmentation CNNs. Furthermore, the proposed 3D three-pass tracking systems outperforms the tracking system available in PhagoSight. In addition, the proposed neutrophil selection module allows our system to raise the precision up to 95%, which is necessary for the posterior biological analyses to be meaningful.

Moreover, we have tested the proposed feature extraction system on the characterization of cell migration, to gain insight into those biological processes where migration plays a relevant role. To the best of our knowledge, this is the first time that such a comprehensive automatic analysis of *in vivo* neutrophil migration has been performed. The total population have reached hundreds of neutrophils with thousands of time instances. We have shown that we can produce an interpretable description of the short-term cell behaviors, for cells belonging to four biological populations. To this end, we have proposed a hierarchical explainability method that reveals the more prominent features for characterizing each behavior or group of behaviors, general enough to be applied to different scenarios. In particular, although neutrophil migration description is difficult (behaviors are similar to each other), the conclusions reached by our method, based on an unbiased analysis over a large neutrophil population, support the biological hypotheses under study, namely that inflammation can be described by the individual behavior of its cellular constituents.

The main lines of further research comprise the consideration of the temporal dimension in the segmentation task, either through 4D supervision of 3D segmentation CNN, or through a full 4D architecture. In this way, cell detection will benefit from temporal continuity (cell instances overlap between two adjacent time steps) and even cell collision can be detected. With respect to the hierarchical explainability method, the features extracted can be refined experimentally, and more complex hierarchical approaches can be explored. Finally, Latent Dirichlet Allocation models (Blei et al., 2003), successfully applied to other biomedical fields, e.g. providing interpretable biological information for gene expression in single-cell RNA sequencing (Wu et al., 2021), could be employed for behavioral cell profiling in our task.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Miguel Molina-Moreno: Conceptualization, Methodology, Software, Investigation, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Iván González-Díaz:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition. **Jon Sicilia:** Investigation, Resources, Visualization. **Georgiana Crainiciuc:** Investigation, Resources, Data curation. **Miguel Palomino-Segura:** Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Andrés Hidalgo:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition. **Fernando Díaz-de-María:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition.

Acknowledgments

This work has been partially supported by the National Grant TEC2017-84395-P of the Spanish Ministry of Economy and Competitiveness, Madrid Regional Government and Universidad Carlos III de Madrid through the project SHARON-CM-UC3M, RTI2018-095497-B-I00 from Ministerio de Ciencia e Innovación (MICINN) and HR17_00527 from Fundación La Caixa to A.H. M.M-M. is supported by the Spanish Ministry of Education, Culture and Sports FPU Grant FPU18/02825. M.P-S. is supported by a Federation of European Biochemical Societies long-term fellowship. J.S. is supported by a fellowship (PRE2019-089130) from MICINN.

References

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S., 2017. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 33, 2424–2426. doi:10.1093/bioinformatics/btx180.

Ballesteros, I., Rubio-Ponce, A., Genua, M., et al., 2020. Co-option of neutrophil fates by tissue environments. Cell 183, 1282–1297. doi:10.1016/j.cell.2020.10.003.

Barabino, G.A., Platt, M.O., Kaul, D.K., 2010. Sickle cell biomechanics. Annu. Rev. Biomed. Eng. 12, 345–367. doi:10.1146/annurev-bioeng-070909-105339.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth and Brooks.

Çiçek, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Springer International Publishing, Cham, pp. 424–432.

Chen, L. C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Chen, S., Ma, K., Zheng, Y., 2019. Med3D: transfer learning for 3D medical image analysis. CoRR. abs/1904.00625.

Chen, W., Wei, H., Peng, S., Sun, J., Qiao, X., Liu, B., 2019. HSN: hybrid segmentation network for small cell lung cancer segmentation. IEEE Access 7, 75591–75603. doi:10.1109/ACCESS.2019.2921434.

Cheng, Y.-H., Lin, T.-C., Ding, J.-J., Wu, Y.-F., Lin, S.-J., 2017. Adaptive 3D cell segmentation and tracing algorithm using convex separation and histogram information for vivo images. In: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 133–138. doi:10.1109/ICMEW.2017.8026264.

Choi, D., 2020. Pytorch deeplabv3+ 3D. https://github.com/ChoiDM/pytorch-deeplabv3plus-3D.

Combs, C.A., Shroff, H., 2017. Fluorescence microscopy: a concise guide to current imaging methods. Curr. Protoc. Neurosci. 79, 2.1.1–2.1.25. doi:10.1002/cpns.29.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1, 224–227. doi:10.1109/TPAMI.1979.4766909.

Dhawan, A.P., D'Alessandro, B., Fu, X., 2010. Optical imaging modalities for biomedical applications. IEEE Rev. Biomed. Eng. 3, 69–92. doi:10.1109/RBME.2010.2081975.

Di Pilato, M., Kfuri-Rubens, R., Pruessmann, J.N., et al., 2021. CXCR6 positions cytotoxic T-cells to receive critical survival signals in the tumor microenvironment. Cell 184, 4512–4530.e22. doi:10.1016/j.cell.2021.07.015.

Di Pilato, M., Palomino-Segura, M., Mejías-Pérez, E., et al., 2021. Neutrophil subtypes shape HIV-specific CD8 T-cell responses after vaccinia virus infection. npj Vaccines 6. doi:10.1038/s41541-021-00314-7.

Dufour, A., Thibeaux, R., Labruyere, E., Guillen, N., Olivo-Marin, J., 2011. 3-D active meshes: Fast discrete deformable models for cell tracking in 3-D time-lapse microscopy. IEEE Trans. Image Process. 20, 1925–1937. doi:10.1109/TIP.2010.2099125.

Fazli, M.S., Vella, S.A., Moreno, S.N.J., Ward, G.E., Quinn, S.P., 2018. Toward simple & scalable 3D cell tracking. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 3217–3225. doi:10.1109/BigData.2018.8622403.

Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378–382. doi:10.1037/h0031619.

García-Prieto, J., Villena-Gutiérrez, R., Gómez, M., Bernardo, E., Pun-García, A., García-Lunar, I., Crainiciuc, G., Fernández-Jiménez, R., Sreeramkumar, V., Bourio-Martínez, R., García-Ruiz, J.M., Del Valle, A.S., Sanz-Rosa, D., Pizarro, G., Fernández-Ortiz, A., Hidalgo, A., Fuster, V., Ibanez, B., 2017. Neutrophil stunning by metoprolol reduces infarct size. Nat. Commun. 8, 14780. doi:10.1038/ncomms14780.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proc. AISTATS 249–256.

Harder, N., Bodnar, M., Eils, R., Spector, D.L., Rohr, K., 2011. 3D segmentation and quantification of mouse embryonic stem cells in fluorescence microscopy images. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 216–219. doi:10.1109/ISBI.2011.5872391.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN. CoRR. abs/1703.06870.

Henry, K.M., Pase, L., Ramos-Lopez, C.F., Lieschke, G.J., Renshaw, S.A., Reyes-Aldasoro, C.C., 2013. PhagoSight an open-source MATLAB ® package for the analysis of fluorescent neutrophil and macrophage migration in a zebrafish model. PLoS ONE 8, 1–11. doi:10.1371/journal.pone.0072636.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844. doi:10.1109/34.709601.

Inman, H.F., Bradley Jr., E.L., 1989. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. Commun. Stat. - Theory Methods 18, 3851–3874. doi:10.1080/03610928908830127.

Jaeger, P.F., Kohl, S.A.A., Bickelhaupt, S., Isensee, F., Kuder, T.A., Schlemmer, H., Maier-Hein, K.H., 2018. Retina U-Net: embarrassingly simple exploitation of segmentation supervision for medical object detection. CoRR. abs/1811.08661.

Jensen, E.C., 2012. Types of imaging, Part 2: an overview of fluorescence microscopy. Anat. Rec. 295, 1621–1627. doi:10.1002/ar.22548.

Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2016. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. CoRR. arXiv preprint arXiv:1603.05959.

Kong, J., Wang, F., Teodoro, G., Liang, Y., Zhu, Y., Tucker-Burden, C., Brat, D.J., 2015. Automated cell segmentation with 3D fluorescence microscopy images. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 1212–1215. doi:10.1109/ISBI.2015.7164091.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 1097–1105.

Lefevre, J.G., Koh, Y.W.H., Wall, A.A., Condon, N.D., Stow, J.L., Hamilton, N.A., 2020. LLAMA: a robust and scalable machine learning pipeline for analysis of cell surface projections in large scale 4D microscopy data. bioRxiv doi:10.1101/2020.12.10.420414.

Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007. doi:10.1109/ICCV.2017.324.

Lloyd, S.P., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory 28, 129–137.

Milletari, F., Navab, N., Ahmadi, S., 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. doi:10.1109/3DV.2016.79.

Orth, A., Schaak, D., Schonbrun, E., 2017. Microscopy, meet big data. Cell Syst. 4, 260–261. doi:10.1016/j.cels.2017.03.009. http://www.sciencedirect.com/science/article/pii/S2405471217300947.

Padfield, D., Rittscher, J., Roysam, B., 2008. Spatio-temporal cell segmentation and tracking for automated screening. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 376–379. doi:10.1109/ISBI.2008.4541011.

Pécot, T., Bouthemy, P., Boulanger, J., Chessel, A., Bardin, S., Salamero, J., Kervrann, C., 2015. Background fluorescence estimation and vesicle segmentation in live cell imaging with conditional random fields. IEEE Trans. Image Process. 24, 667–680. doi:10.1109/TIP.2014.2380178.

Pijuan, J., Barceló, C., Moreno, D.F., Maiques, O., Sisó, P., Marti, R.M., Maciá, A., Panosa, A., 2019. In vitro cell migration, invasion, and adhesion assays: from cell imaging to data analysis. Front. Cell Dev. Biol. 7, 107. doi:10.3389/fcell.2019.00107.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 91–99.

Richards, R., Mason, D., Lévy, R., Bearon, R., Sée, V., 2018. 4D imaging and analysis of multicellular tumour spheroid cell migration and invasion. bioRxiv doi:10.1101/443648.

Schindelin, J., Arganda-Carreras, I., Frise, E.e.a., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. doi:10.1038/nmeth.2019.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 640–651. doi:10.1109/TPAMI.2016.2572683.

Sreeramkumar, V., Adrover, J.M., Ballesteros, I., Cuartero, M.I., Rossaint, J., Bilbao, I., Nácher, M., Pitaval, C., Radovanovic, I., Fukui, Y., McEver, R.P., Filippi, M.D., Lizasoain, I., Ruiz-Cabello, J., Zarbock, A., Moro, M.A., Hidalgo, A., 2014. Neutrophils scan for activated platelets to initiate inflammation. Science 346, 1234–1238. doi:10.1126/science.1256478.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. doi:10.1109/CVPR.2015.7298594.

Tokuoka, Y., Yamada, T.G., Mashiko, D., Ikeda, Z., Hiroi, N.F., Kobayashi, T.J., Yamagata, K., Funahashi, A., 2020. 3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis. npj Syst. Biol. Appl. 6. doi:10.1038/s41540-020-00152-8.

van der Maaten, L., Hinton, G., 2008. Visualizing high-dimensional data using t-SNE.

Welch, G., Bishop, G., 1995. An Introduction to the Kalman Filter. Technical Report. USA.

Wu, X., Wu, H., Wu, Z., 2021. Penalized latent Dirichlet allocation model in single-cell RNA sequencing. Stat. Biosci. doi:10.1007/s12561-021-09304-8.