3-D Quantification of Filopodia in Motile Cancer Cells

Carlos Castilla, *Student Member, IEEE*, Martin Maška, *Member, IEEE*, Dmitry V. Sorokin, Erik Meijering, *Senior Member, IEEE*, and Carlos Ortiz-de-Solórzano, *Senior Member, IEEE*

Abstract—We present a 3D bioimage analysis workflow to quantitatively analyze single, actin-stained cells with filopodial protrusions of diverse structural and temporal attributes, such as number, length, thickness, level of branching, and lifetime, in time-lapse confocal microscopy image data. Our workflow makes use of convolutional neural networks trained using real as well as synthetic image data, to segment the cell volumes with highly heterogeneous fluorescence intensity levels and to detect individual filopodial protrusions, followed by a constrained nearest-neighbor tracking algorithm to obtain valuable information about the spatio-temporal evolution of individual filopodia. We validated the workflow using real and synthetic 3-D time-lapse sequences of lung adenocarcinoma cells of three morphologically distinct filopodial phenotypes and show that it achieves reliable segmentation and tracking performance, providing a robust, reproducible and less time-consuming alternative to manual analysis of the 3D+t image data.

Index Terms—Filopodium segmentation and tracking, actin cytoskeleton, confocal microscopy, 3D skeletonization, Chan-Vese model, convolutional neural network, deep learning.

I. INTRODUCTION

CELL migration is an essential phenomenon that underlies many physiological and pathological processes, such as

wound healing, embryonic development, or the formation of cancer metastases, with a critical importance attributed to filopodia [1]–[3]. These are cell membrane protrusions composed of dynamic bundles of aligned actin filaments that act as guidance sensors and promoters of the locomotion activity [4]–[6].

The state-of-the-art optical microscopy and fluorescent reporters facilitate the observation of filopodial formation and dynamcs in three-dimensional microenvironments at unprecedented spatio-temporal resolution [7]. However, a comprehensive understanding of the biology of filopodia has yet to be achieved [8], [9]. This predominantly stems from the lack of robust automatic bioimage analysis workflows for the quantitative analysis of filopodium-mediated processes in 3D+t image data, which impels biologists to analyze only pure 2D+t image data or maximum intensity projections of acquired 3D+t image data [8]–[11]. Indeed, a properly designed 3D workflow would avoid the subjectivity, high error-proneness, and extreme labor involved in manually analyzing large amounts of 3D+t image data [12]. Furthermore, a fully 3D analysis would provide biological results of much higher significance than those provided by 2D approaches, by analyzing the role of arbitrarily oriented filopodial protrusions in situations of high physiological relevance.

To the best of our knowledge, the existing methods designed to quantitatively analyze cells with protruding filopodia focus mostly on the 2D+t case, exploiting simple intensity-based thresholding algorithms [8]–[11], [13], [14]. Those approaches can barely be applied in 3D+t, since 3D confocal stacks suffer from increasing fluorescence attenuation in the axial direction, display high Poisson noise levels, and are blurred by the axially variant point spread function of the imaging system [15]. Furthermore, the actin-specific labeling, required for a good definition of the filopodial structures, produces heterogeneous labeling of the cell volume that concentrates the fluorescent signal in the cortical cytoskeleton of the cell and leaves the inner parts of the cell practically unstained. We partially addressed these difficulties by applying, slice by slice, a sequence of morphological filters to the negative response of a Hessian-based ridge detector [16]. Apart from only partially exploiting the image information in the axial direction, the single-scale nature of the ridge detector precludes an accurate segmentation of branching filopodia, the thickness of which decreases with increasing levels of branching, due to its limited focus on a single, preset filopodium diameter.
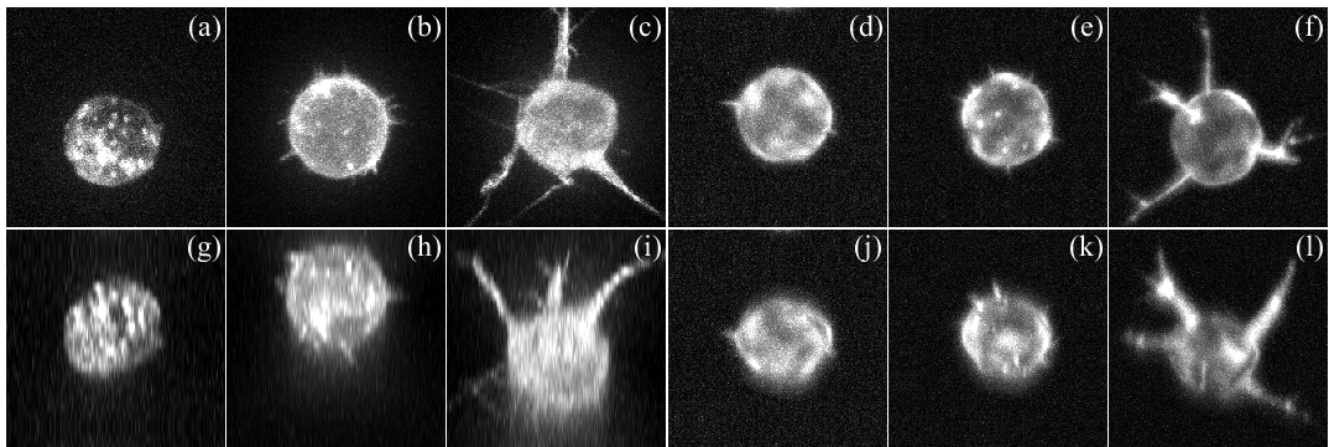
Fig. 1.   Contrast enhanced, maximum intensity projections along the axial (a-f) and tangential direction (g-l) of a WT_R (a,g), OE_R (b,h), PD_R (c,i), WT_S (d,j), OE_S (e,k) and PD_S (f,l) cell containing image stack. Notice the difference in length and number of protrusions of the three different phenotypes in both real (first three columns) and synthetic (last three columns) image data.

In this paper, we present and evaluate a full 3D approach to quantitatively analyze single, actin-stained cells with filopodial protrusions. The proposed approach considerably extends our recent, segmentation-centric routine [17], by integrating it with the quantification and tracking of filopodia. For cell segmentation, our approach exploits deep learning, which has been shown to outperform traditional approaches in many cell tracking [18] and particle tracking [19] tasks. To take full advantage of its learning capabilities, we train a 3D convolutional neural network (CNN) [20] using both real and synthetic image data. The real image data consists of 3D time-lapse videos of actin-stained lung cancer cells of three distinct filopodial phenotypes, segmented using a full 3D segmentation routine built on machine-learning-free concepts, from our previous work on the segmentation of filopodia through a steerable ridge detector [16] and the segmentation of cells through the minimization of the Chan-Vese model (CVS) [21]. The synthetic image data consists of simulations of 3D time-lapse sequences of single motile cells with filopodial protrusions of user-defined structural and temporal attributes, such as the number, thickness, length, level of branching, and lifetime of filopodia [22], [23]. To characterize and track filopodia in time, the cell segmentation masks provided by the CNN are skeletonized and the filopodial tips are subsequently linked into tracks using a constrained nearest-neighbor particle tracker [10]. We demonstrate that our proposed CNN-based approach outperforms our baseline CVS method both in regard to the segmentation of protruding cells, and to the detection and tracking of filopodia. This first attempt at a full 3D analysis of the dynamics of cell protrusions opens the door to the study of the role of actin fibers in homeostasis and cell migration, in highly physiological situations.

The paper is organized as follows: the image datasets used and analyzed in our experiments are briefly described in Section II. The proposed bioimage analysis workflow is explained in detail in Section III and experimentally validated in Section IV. Finally, a discussion of the results, concluding remarks, and suggestions for future work are given in Section V.

II. IMAGE DATA

We used both real and synthetic 3D time-lapse sequences of single lung adenocarcinoma cells embedded in a 3D scaffold made of a mixture of Matrigel and fibrillar collagen. All datasets are available for download through a IEEE Data Port.[1]

A. Real Image Data

The real image data consisted of nine 3D+t image sequences of cells of the A549 lung adenocarcinoma cancer cell line, displaying three different phenotypes of CRMP-2, a protein involved in the formation of microtubules, whose dynamic scaffolding affect the structure and organization of the actin cytoskeleton. Specifically, we used three sequences of wild-type (WT_R) cells (Fig. 1a, 1g), three sequences of cells over-expressing CRMP-2 (OE_R) (Fig. 1b, 1h) and three sequences of cells with a phospho-defective form of CRMP-2 (PD_R) (Fig. 1c, 1i). The corresponding phenotypes are shown in Fig. 1: WT_R cells display very few, short filopodia; OE_R cells display a large number of slightly longer filopodia, compared to the WT_R cells; and PD_R cells display aberrantly long and thick branching filopodia. The cells were transiently transfected with LifeAct-GFP to visualize the actin protein conjugated with the Green Fluorescent Protein at the location of actin fibers. All videos were acquired on a Ultraview ERS (Perkin Elmer, Inc., Waltham, MA, USA) spinning disk confocal microscope, exciting with the 488 nm line of an Ar/Kr laser, through a Plan-Apochromatic $63\times$ 1.20 NA water immersion objective lens (Carl Zeiss, AG., Wetzlar, Germany). The videos contained one cell, imaged every two minutes during one hour. The original image voxel size of $0.126 \times 0.126 \times 1.0$ μm was resampled in the axial direction using cubic spline interpolation to obtain isotropic image data with a voxel size of $0.126 \times 0.126 \times 0.126$ μm.
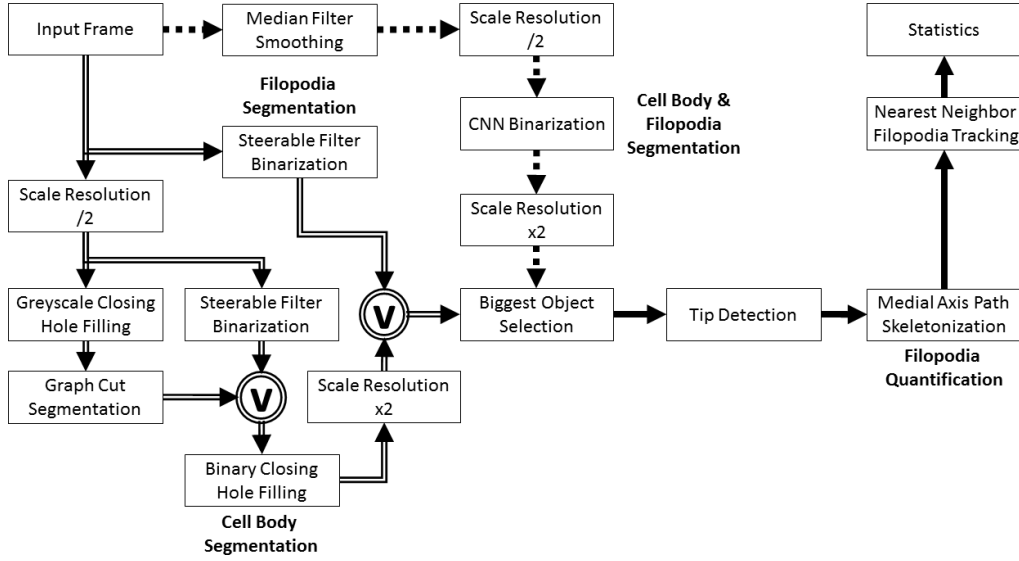
Fig. 2.   Analysis workflow: CVS (double lines), CNN (square dotted lines) and common (solid lines) steps. The V inside the circle indicates the use of the binary maximum operator.

B. Simulated Image Data

The synthetic data was created using our FiloGen simulator [23], with the aim of qualitatively and quantitatively resembling the real time-lapse sequences described in the previous section, in terms of cell image properties (intensity, texture, noise) as well as filopodial tip dynamics. To this end, nine sequences of single cells with the three mentioned CRMP-2 phenotypes: wild-type (WT_S), over-expressing (OE_S) and phospho-defective (PD_S) were computer-generated, having the same isotropic spatial and temporal resolution as the real image data (Fig. 1.d-f, Fig. 1.j-l). These sequences were used both to evaluate the segmentation and tracking methods and to enrich the training sets of the CNNs used to segment cells and detect filopodial tips.

C. Ground Truth

We use an extensive *ground truth*, for training and evaluation purposes. The ground truth for the cell segmentation of the real sequences (GTS_R) consisted of manual annotations of 10 randomly selected slices, from three frames of each sequence. Following the same procedure used in the Cell Tracking Challenge [18], the annotation was done by three experts, who manually delineated the cell contours in the selected slices. The three annotations were merged into one final consensus annotation using a majority voting approach. The ground truth for the detection of filopodial tips in the real sequences (GTD_R) consisted of a single manual annotation of the voxel coordinates of the filopodial tips. The annotation was done in three frames of each of the nine sequences.

The ground truth for the tracking of filopodial tips in the real sequences (GTT_R) consisted of manual annotations of the spatio-temporal positions (i.e. tracks) of the tips of clearly established filopodial protrusions. These annotations were performed by three experts, and later combined in one final ground truth by maximizing the overall Linear Oriented

Forests Matching measure [24] between the combined reference annotation and each of the three manual annotations.

The synthetic GT consisted of the inherently provided cell segmentation masks (GTS_S) used to produce the synthetic datasets, the spatial coordinates of all filopodial tip detections (GTD_S), and their tracks (GTT_S), in all frames of the nine synthetic sequences.

III. Methods

A. Baseline Segmentation Approach

Our baseline cell segmentation approach is based on the minimization of the Chan-Vese model (CVS). This method was used to produce training pairs for the CNN-based segmentation method described in Section III-C, and as a reference to calculate the improvement introduced by the CNN, in regard to the segmentation, filopodial tip detection and tracking accuracy. The workflow of the baseline approach, which separately addresses the segmentation of the cell body and the filopodia to later combine both, is shown in Fig. 2 (double line tracks) and described next.

To segment the inner part of the cell body in each image frame of the sequence, we first downsample the frame by a factor of two, to improve computational efficiency. Then, we apply a grayscale closing with a spherical structuring element of radius C_{gc} followed by a hole filling operation that fills the interior of the cell, left unstained by the actin-specific staining that concentrates at the cortical regions of the cell. Then, we segment the resulting image by minimizing the Chan–Vese model using graph cuts (GC) [21]. Namely, given an input scalar image $u \colon \Omega \to \mathbb{R}$ defined over a d–dimensional image domain $\Omega \subset \mathbb{R}^d$, the model aims at partitioning the image domain into two possibly disconnected regions Ω_1 (foreground) and Ω_2 (background) of minimum intra-region variance, separated by a closed contour $C(\Omega = \Omega_1 \cup \Omega_2 \cup C)$,

by minimizing

$$E_{CV}\left(C, c_1, c_2\right) = \mu \, |C| + \gamma_1 \int_{\Omega_1} \left(u\left(x\right) - c_1\right)^2 dx$$
$$+ \gamma_2 \int_{\Omega_2} \left(u\left(x\right) - c_2\right)^2 dx$$

where $|\bullet|$ stands for the contour length, μ, γ_1 and γ_2 are the curvature, foreground and background weights. c_1 and c_2 are the average intensity levels inside Ω_1 and Ω_2, respectively.

The cortical, actin-labeled regions of the cell body are segmented by thresholding the downsampled grayscale image after applying a second-order steerable filter [25], which targets arbitrarily oriented ridges. Let $*$ denote the convolution operator, G_σ be the Gaussian kernel with the standard deviation σ_c, and H_u be the Hessian matrix operator applied to the image u. For every grid point $x \in \Omega$, the steerable filter response $\mathcal{F}_\sigma\left(u\left(x\right)\right)$ is

$$\mathcal{F}_\sigma\left(u\left(x\right)\right) = \lambda_D$$

where λ_D is the dominant eigenvalue of $H_{G\sigma*u}(x)$. As the filter response is negative in the foreground areas, the binary mask $\mathbb{S}$ is

$$\mathbb{S}\left(x\right) = \begin{cases} 1, & \text{if } F_\sigma\left(u\left(x\right)\right) < T \\ 0, & \text{otherwise} \end{cases}$$

with T being a threshold (T_c).

The inner part and cortical regions of the cell body are combined using a binary maximum operator, followed by binary closing with a spherical structuring element of radius C_{bc} and by hole filling. The resulting cell body mask is then up-sampled back to its original resolution.

To segment filopodia we threshold the original image, preprocessed using a second order steerable filter with the standard deviation σ_f and threshold T_f. This filter enhances slim protrusions that may be missed in the previous steps and more accurately highlights the contour of the cell. This result is combined with the cell body mask using a binary maximum operator, producing the final cell segmentation mask for the CVS method.

We fix the value of γ_2 to 1. All other parameters (μ, γ_1, C_{gc}, σ_c, T_c, C_{bc}, σ_f, T_f) are specifically optimized for each phenotype using a genetic algorithm (GA) [26] that maximizes the Jaccard coefficient (JC), which measures the similarity between the segmentation output and the corresponding ground truth (GTS_R, GTS_S). The parameter search ranges, and the parameters that led to the optimal JC values are listed in Supplementary Tables I & II. A parameter sensitivity analysis is provided in Supplementary Note 1.[2]

B. CNN-Based Segmentation

We use a 3D CNN to segment the entire cell -i.e. cell body and filopodia-, following the workflow shown in Fig. 2 (square dotted line tracks). The CNN structure, training and classification strategies have been described previously [17].

Briefly, we use a batch-normalized version of the 3D U-Net [20] made of four layers with 32-64, 64-128, 128-256 and 256-512 feature maps in the convolution path and three layers with 512-256-256, 256-128-128, 128-64-64 feature maps in the up-convolution path. We modify the data augmentation layer by inserting a batch size of one with a window size of $132 \times 132 \times 124$ voxels, and by increasing both the random elastic grid spacing and deformation magnitude to 80 and 70 voxels respectively in all axes to allow realistic deformations of our large protruding cells. The random offsets for the data augmentation layer are set to fit the size of our image data. A complete set of 3D rotations was performed to cover all possible orientations of the cells. Before training and classification, we apply a median filter within a neighborhood of $3 \times 3 \times 3$ voxels and down-sample the frames by a factor of two, to improve computation efficiency and memory usage while allowing a large enough cell analysis context.

To train the CNN we use both real and synthetic 3D image frames, along with their corresponding ground truths. For the real frames, we use GTS_R along with the segmentation masks obtained using the CVS method for all the remaining slices not included in GTS_R. Note that the CNN is trained simultaneously for all cell phenotypes. For the synthetic sequences, we use GTS_S. The ground truth components are differently weighted to improve the training of the network, reducing undersegmentation or oversegmentation of the cell body and filopodia. Namely, in GTS_R, we use a weight of 3 for the background and 30 for the foreground cell voxels. In the 3D frames segmented using the CVS method, we use a weight of 1 for both background and foreground voxels. Finally, in GTS_S, the background is assigned a weight of 3 and the cell body and filopodia are given a higher weight of 60. We also add a stronger weight of 300 to the voxels located at the filopodial tips to force the CNN to refine the segmentation in those critical areas. Each weight was obtained following an initial manual half-interval search within a large [0,1000] interval, by visually inspecting the output of the CNN after 2000 iterations. Once the weights had been properly set, a complete training phase of 100000 iterations was launched. The main parameters used to train the CNN are provided in Supplementary Note 2.[3]

Each down-sampled image is classified using an analysis window of $132 \times 132 \times 124$ voxels. As our images are larger than this window, each image is divided into a set of tiles that cover the entire image volume, leaving the overlap of $44 \times 44 \times 44$ voxels between tiles. For each tile, the CNN provides two output probability maps (background and cell). To reduce artifacts near the boundaries of the analysis window, we average -separately for each class- four weighted probability maps corresponding to the original and three transformed versions of the image: a mirrored version relative to the three axes, a translated version in the three directions and a mirrored and translated version relative to the three axes. Before averaging, the probability maps are weighted using a context window of the same size as the

---

[2]Supplementary materials are available in the supplementary files/multimedia tab

[3]Supplementary materials are available in the supplementary files/multimedia tab

analysis window, with weights corresponding to the Euclidean distance of each voxel to the border of the context window. This weighting strategy assigns more relevance to the voxels located closer to the center of the analysis window than to those located near the borders. After that, we obtain a weighted probability map for each class, corresponding to the entire 3D image. To obtain the final voxel values (0 for background and 1 for cell) we apply the soft max transfer function to the probability maps. This produces the final cell binary segmentation mask which is then up-sampled back to the original resolution.

C. Filopodial Tip Detector

To detect the tips of the filopodia from the 3D cell binary segmentation masks, we use a 3D CNN (Fig. 3e) with the same architecture as the one used to segment the cells. The random elastic grid spacing is set to 125 voxels and the random elastic deformation magnitude to 80. All other parameters remain the same.

The binary frames are down-sampled to a third of their size for training and classification, to improve computation efficiency and reduce image noise while allowing for a large enough analysis context. The CNN is trained with binary masks obtained using the baseline method (real sequences), and the GTS_S (synthetic sequences), along with their corresponding ground truth annotations GTD_R and GTD_S. Specifically, we used a subset of GTD_R and GTD_S: three frames from two real videos per phenotype and three frames from two synthetic videos of the OE_S and PD_S phenotypes.

To provide a large detection context, the voxel corresponding to each annotated coordinate is dilated with a sphere of four voxels in radius, and the voxels intersecting with the boundary of the filopodia are labeled as a tip. A strong weight of 15 is assigned to the tips and their 1-voxel 8-connected neighborhood. Voxels located two voxels from the tips, in terms of the city-block distance, are assigned a weight of 3. The rest of the cell body is given a weight of 2 and the background a weight of 1 (Supplementary Figure 1).[4] All weights are fixed visually after a half-interval search, after 2000 CNN training iterations, within a [0,1000] search interval.

The CNN classifies voxels of the 3D cell binary masks as tips (1) or background (0). We use the same analysis window as in the segmentation CNN. In order to reduce false negative detections near the context window boundaries, we apply the same translations and mirrored transformations described for the segmentation CNN. Finally, to merge detections that are close to one another, we apply a morphological closing with the spherical structuring element of two voxels in radius.

D. Filopodial Tip Refinement

We refine the size and location of the tips detected by the CNN. To this end we calculate the Euclidean distance transform and the quasi-Euclidean geodesic distance transform

---

[4]Supplementary materials are available in the supplementary files /multimedia tab

of the binary cell segmentation masks. Then each connected component representing a detected tip is replaced by a one-voxel representative, defined as the voxel of the connected component located at a minimum Euclidean distance from the cell boundary. Then, these representative voxels are moved along their nearest maximum geodesic paths, from their initial locations to the closest local maxima of the geodesic distance transform, calculated from the center of the cell binary mask, defined as the global maximum in the Euclidean distance transform (Fig. 3c,3f).

E. Cell Skeletonization

We next trace the path from each detected tip to the center of the cell, following the medial axis path. To this end we start from the voxel that contains the tip, and follow the maximum of the Euclidean distance towards the cell center, selecting only voxels with lower geodesic distance than the last voxel included in the current path (Fig. 3d,3g). The algorithm stops when the path reaches the cell center, or when it converges to a previously computed path.

F. Filopodial Tip Tracking

To track the tips of the filopodia, we use a 3D extension of the constrained Euclidean-distance based nearest-neighbor particle tracker [10]. This greedy linking algorithm associates filopodial tips between two consecutive frames by minimizing the total gated Euclidean distance (gating distance = 40 voxels) between the two corresponding sets of filopodia. The filopodial base width, defined as the minimum Euclidean distance from a point in the cell skeleton, to any boundary point, as well as the minimum filopodial length, defined as the Euclidean distance from its tip to its base following the skeleton path, are optimized using the tracking ground truth (i.e., GTT_R, GTT_S). Note that to account for minor shifts in the location of the cells between consecutive frames due to migration or small drifting of the microscope stage, we calculate the position of the centroid of the cell binary masks in all frames, and correct the tip coordinates before tracking, by applying the corresponding translation transformation.

G. Implementation

All software development was done on a workstation with the following specifications: Intel Xeon E5-2637 v4 3.5 GHz * 2, 64 GB RAM, Nvidia Titan X (Pascal) 12 GB RAM, Windows 7 Professional. The CVS preprocessing and segmentation routines, along with the GA used in their parameter optimization phases were implemented in Matlab 2017a (Mathworks, Inc., Natick, MA, USA) with mex calls to our own pre-compiled C++ code for steerable filtering and the Chan-Vese model minimization. The CNNs were implemented using an in-house developed Windows port of the 3D U-Net Caffe [17]. The filopodial tip detection, optimization and skeletonization routines were developed in Matlab. For the filopodial tip tracking routines, we extended to 3D the algorithms and most internal structures of the 2D filopodia quantification software CellGeo [10]. The entire segmentation
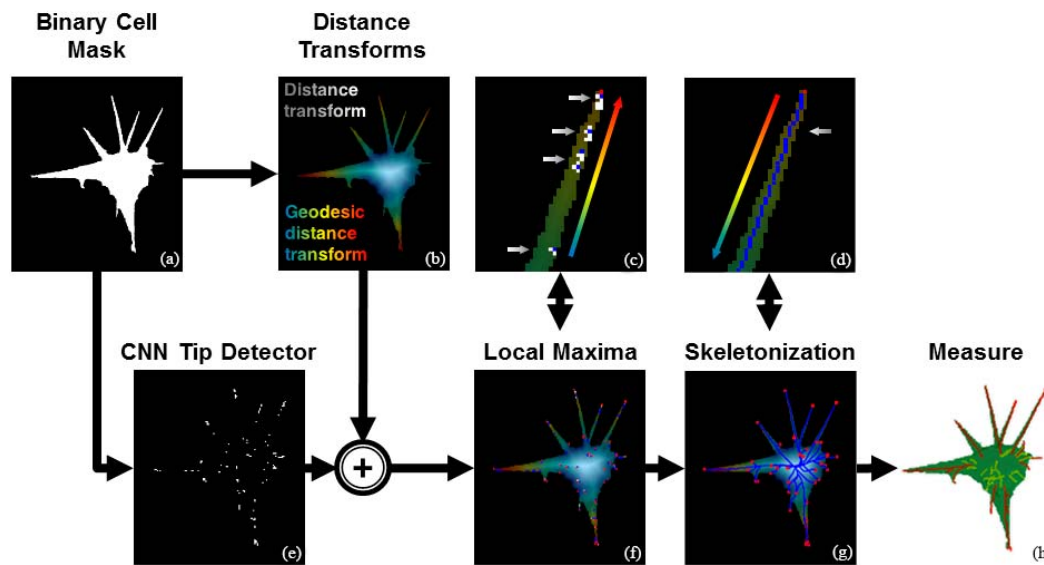
Fig. 3. Filopodial tip detection and skeletonization process. Filopodial tips are detected using the CNN (e) trained with the binary segmentation masks (a) and manual annotations. Using both the distance transform and geodesic distance transform (b), the detected tips shrink to one voxel (blue voxels in (c)) by means of the distance transform (white arrows in (c)) and moved towards the terminal parts of the filopodia (f), following the direction to the local maximum in the geodesic distance transform (multicolored arrows in (c)). The skeleton of the cell (g) is calculated from the filopodial tips following the direction indicated by the geodesic distances (multicolored arrows in (d)) on top of the medial axis path by means of the distance transform (white arrows in (d)). Maximum intensity projections along the axial direction shown.

and tracking pipeline can be run using the scripts provided through the associated IEEE Data Port.[5]

IV. EXPERIMENTAL RESULTS

A. Evaluation Measures

To evaluate the segmentation accuracy, we used the Jaccard coefficient. For a given ground truth X and segmentation output Y, $JC = |X \cap Y| / |X \cup Y|$. To evaluate the filopodial tip detection and tracking accuracy, we used the Linear Oriented Forest Matching measure [24] (i.e., $LOFM_D$ for detection accuracy, $LOFM_R$ for recall, $LOFM_L$ for linking accuracy, and RMSE for localization accuracy), using the default weight configuration ($w_{FN} = 1$, $w_{FP} = 1$, $w_{EA} = 1.5$, $w_{ED} = 1$) and a gating distance ε. All measures fall in the [0, 1] interval, with higher values corresponding to higher accuracy, except RMSE (in μm) falls in the [0, ε) interval, where lower values correspond to higher localization accuracy.

B. Segmentation Performance

The segmentation methods were evaluated on both our real and synthetic datasets using a *3-fold* cross-validation strategy. To this end, the ground truth (GTS_R and GTS_S) was split in a two-to-one basis. Two videos per phenotype were used to optimize (CVS) or train (CNN) the method, and the third video was used to evaluate the segmentation. The CNN method was trained to classify any cell, regardless of its phenotype or data type (real or synthetic), while the CVS method required specific optimization for each phenotype and data type to achieve usable results. Fig. 4 shows representative segmentation results of the CVS method. Fig. 5 shows a qualitative comparison of both CVS and CNN methods. Maximum intensity projection (MIP) videos of the complete set of results are provided

through the associated IEEE Data Port.[6] The comparative quantitative evaluation of the segmentation results obtained using the CVS and CNN segmentation methods is shown in Table I.

A five-way ANOVA (factors 1: *video type*; 2: *segmentation method*; 3: *phenotype*; 4: *video sequence*; 5: *video frame*) was performed to test for significant differences in the *video type*, *segmentation method* and *phenotype*. The ANOVA results revealed a global effect of *segmentation method* ($F(107)=6.51$, $p=0.0128$) confirming that the CNN provides globally significantly better results than the CVS. Furthermore, we found interaction of *video type∗segmentation method* ($F(107)=5.88$, $p=0.0178$) indicating that the CNN also outperforms the CVS specifically on the real datasets. No other interactions were found, indicating that both methods perform similarly when looking individually at each phenotype and data type.

C. Filopodial Tip Detection and Tracking Evaluation

We next evaluated the detection and tracking of filopodial tips extracted from the segmentation masks obtained using either the CVS or CNN methods. This was done using the entire GTT_R and GTT_S. The optimization of the critical parameters of the skeleton that define what is considered a filopodium -critical width (cW_S, cW_R) and minimum length (mL_S, mL_R)- was done differently for the synthetic and real phenotypes. In the synthetic datasets, for which we have absolute ground truth (i.e., we know the exact tip positions of all filopodia in the GTT_S), cW_S was optimized globally for all phenotypes and mL_S was optimized separately for each phenotype using the detection measure $LOFM_D$. A pruning value of 1 μm was applied to the GTT_S, to prevent the
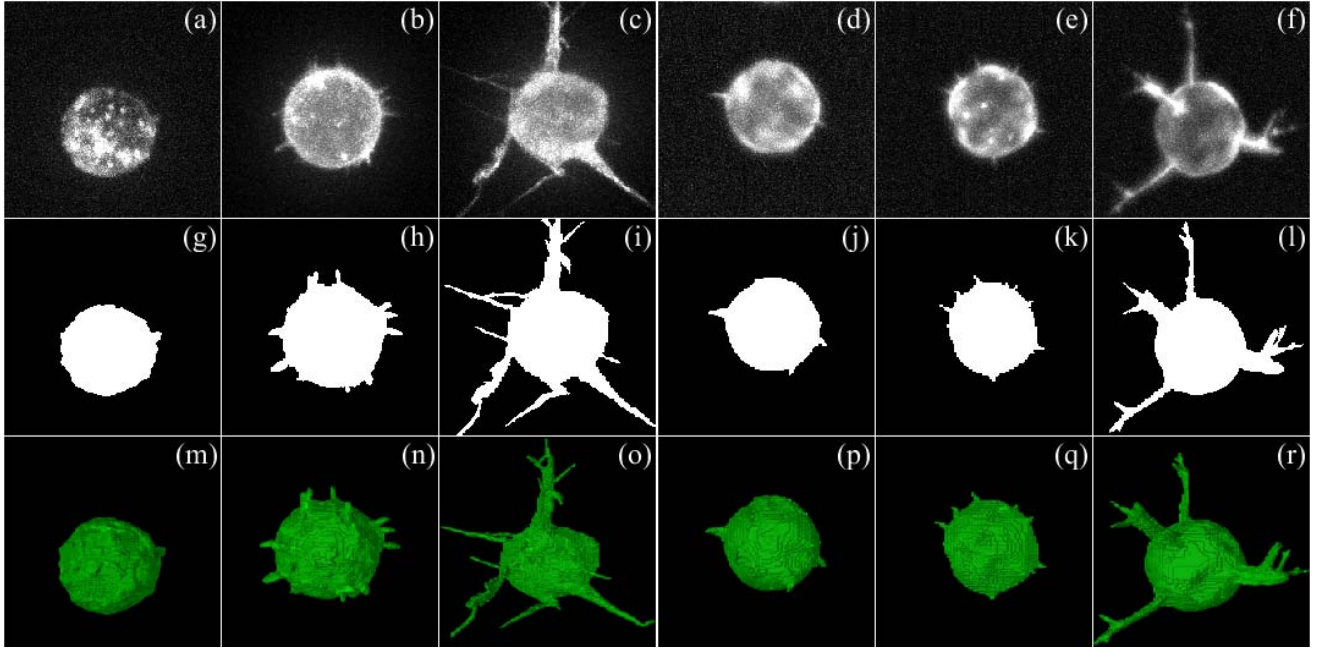
Fig. 4.　Maximum intensity projections along the axial direction of contrast-enhanced grayscale images (a-f) and segmentation masks (g-l) of a (a,g) WT_R, (b,h) OE_R, (c,i) PD_R, (d,j) WT_S (e,k) OE_S and (f,l) PD_S cell. 3D renderings of the segmented cells are shown in the last row (m-r). CVS method results shown.

TABLE I

AVERAGE JACCARD COEFFICIENT – SYNTHETIC AND REAL DATASETS – CVS AND CNN METHODS

Dataset	Synthetic			Real		
Phenotype	Wild-type	Over-expressing	Phospho-defective	Wild-type	Over-expressing	Phospho-defective
CVS	**0.888 (±0.008)**	0.835 (±0.145)	**0.838 (±0.041)**	0.776 (±0.132)	0.858 (±0.176)	0.785 (±0.090)
CNN	0.885 (±0.01)	**0.857 (±0.030)**	0.825 (±0.052)	**0.888 (±0.016)**	**0.915 (±0.021)**	**0.853 (±0.036)**

Average of the Jaccard coefficient of the 3-fold cross validation subsets for CVS and CNN methods. Note that CVS method was optimized separately for each phenotype and dataset type (real and synthetic) but the CNN method was optimized taking into account the three phenotypes and the two dataset types together.

inclusion of detections near or below the resolution in the axial direction. In the real datasets, the ground truth GTT_R is necessarily incomplete (not absolute as in the synthetic datasets), since it is impracticable to visually decide if very short and short-lived detections are filopodia or noise. Therefore, GTT_R contains only clearly established filopodia. Consequently, the critical width (cW_R) was fixed as the average value obtained in the optimization of the synthetic datasets, while the minimum length (mL_R) was optimized per phenotype using the recall measure $LOFM_R$. With this we assume that the algorithm may produce a significant number of detections, and concentrate the validation on the true detections, their linking and localization. The tip association search area or gating distance for the linking algorithm ε, was set based on the median filopodial length per phenotype, calculated from the synthetic data. The optimal values of cW_S, cW_R, mL_S and mL_R and ε for each phenotype are shown in Supplementary Table III.

The filopodial tip detection, recall, localization and tracking measures obtained for the cells segmented using the baseline CVS and CNN methods are shown in Table II, along with the corresponding measures obtained for the synthetic GT segmentation masks (GTS_S), which can be seen as the upper-bound performance that the pipeline can achieve, assuming perfect segmentation masks.

For the detection and recall measures, three four-way ANOVA analyses (factors: 1: *segmentation method*; 2: *phenotype*; 3: *video sequence*; 4: *video frame*) were performed. The ANOVAs revealed a main effect of *segmentation method* globally for the detection measure ($F(467)=28.69$, $p<0.001$), and individually for the synthetic ($F(467)=74.72$, $p<0.001$) and real ($F(539)=113.82$, $p<0.001$) datasets for the recall measure. In all cases, the CNN outperformed the CVS. Particular interactions found by the previous ANOVAs for each *segmentation method* and *phenotype* are shown in Table II.

For localization and linking measures, a four-way ANOVA (factors: 1: *video type*; 2: *segmentation method*; 3: *phenotype*; 4: *video sequence*) was performed. The ANOVA results for the localization measure revealed a main effect of *segmentation method* ($F(35)=15.26$, $p=0.0013$), revealing a significantly better performance of the CNN versus the CVS method. Particular interactions are highlighted in Table II.

The ANOVA results for the linking measure did not reveal any interaction.

TABLE II
TRACKING RESULTS – REAL AND SYNTHETIC DATASETS – CVS AND CNN METHODS

Metric	Phenotype / Method	Synthetic			Real		
		Wild-type	*Over-expressing*	*Phospho-defective*	*Wild-type*	*Over-expressing*	*Phospho-defective*
Detection (LOFM$_D$) [0 1]	CVS	0.660 (±0.412)	0.793 (±0.200)	**0.806* (±0.110)**	N/A	N/A	N/A
	CNN	**0.970* (±0.142)**	**0.905* (±0.204)**	0.647 (±0.141)	N/A	N/A	N/A
	GT Masks	0.987 (±0.022)	0.970 (±0.018)	0.961 (±0.013)	N/A	N/A	N/A
Recall (LOFM$_R$) [0 1]	CVS	0.740 (±0.354)	0.889 (±0.106)	0.827 (±0.111)	0.617 (±0.387)	0.716 (±0.243)	0.764 (±0.219)
	CNN	**0.973* (±0.140)**	**0.968* (±0.065)**	**0.869 (±0.092)**	**0.783* (±0.284)**	**0.838* (±0.188)**	**0.926* (±0.088)**
	GT Masks	0.987 (±0.023)	0.988 (±0.005)	0.961 (±0.013)	N/A	N/A	N/A
Localization (RMSE) [0, ε) *(in μm)*	CVS	0.613 (±0.233)	0.931 (±0.043)	1.868 (±0.224)	1.074 (±0.978)	**1.175 (±0.016)**	3.772 (±0.549)
	CNN	**0.564 (±0.070)**	**0.643 (±0.023)**	**1.408 (±0.093)**	**0.774 (±0.671)**	1.235 (±0.051)	**2.798* (±0.353)**
	GT Masks	0.177 (±0.019)	0.173 (±0.008)	0.291 (±0.015)	N/A	N/A	N/A
Linking (LOFM$_L$) [0 1]	CVS	0.983 (±0.013)	0.997 (±0.001)	0.937 (±0.022)	**0.986 (±0.024)**	0.946 (±0.002)	0.886 (±0.083)
	CNN	**0.990 (±0.006)**	**0.998 (±0.001)**	**0.954 (±0.009)**	0.728 (±0.447)	**0.953 (±0.028)**	**0.945 (±0.034)**
	GT Masks	0.997 (±0.001)	0.999 (±0.000)	0.968 (±0.017)	N/A	N/A	N/A

Detection, recall, localization and linking average measure values obtained from the segmentation masks created using the CVS and CNN method. Results obtained for the synthetic GT masks (GTS$_S$) are also shown as a reference, being the maximum achievable values in case perfect segmentation masks are used. Values with * denotes significant differences between methods.
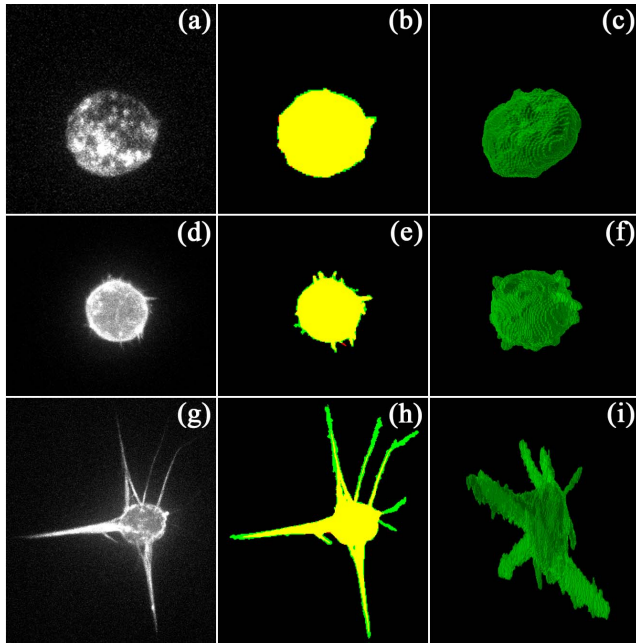

Fig. 5. Maximum intensity projections of the contrast-enhanced grayscale images (left), segmentation masks projections (middle) and CNN rendering results (right) of a WT$_R$ (a-c), OE$_R$ (d-f) and PD$_R$ (g-i) cell. CVS result shown in red, CNN result shown in green. The intersection of both approaches is shown in yellow.

D. Quantification of Filopodial Dynamics

The dynamics of filopodia in the real and synthetic videos segmented using the CNN method, is described in Fig. 6. Namely, for each phenotype we show the average number, average length and lifetime cumulative histograms of the filopodia. In the synthetic datasets, due to the fact that we have the absolute ground truth (GTT$_S$), we performed a two-sample Kolmogorov-Smirnov test to determine if lifetime distributions are equal to those obtained using the GT synthetic masks (5% significance level). The WT$_S$ and OE$_S$ phenotypes satisfy the test while the PD$_S$ does not. This is consistent with the presence of spurious detections in the PD$_S$ phenotype (see Table II), due to the fact that the CNN was optimized to segment the real datasets, not the synthetic datasets. The p-values obtained were 0.9999 for WT$_S$, 0.5309 for OE$_S$ and 0.0001 for PD$_S$.

E. Robustness of the CNN Method

To evaluate the robustness of the originally trained CNN-based method against varying the cellular phenotype, the signal-to-noise ratio (SNR), and the imaging system, represented by its point spread function (PSF) of the microscope, we generated and analyzed new synthetic image datasets. These new datasets –each made of three sequences with the same length and spatio-temporal resolution of the original synthetic datasets- display single cells of two new synthetic geometries (i.e. phenotypes). These phenotypes could be considered intermediate between the OE and PD phenotypes in terms of width and length of their single-branch filopodial protrusions. We named them phospho-over-expressing 1 (PO1$_S$) and phospho-over-expressing 2 (PO2$_S$). Supplementary Table IV shows the filopodial lengths of the original and new synthetic phenotypes. The fluorescence level of the new cell geometries used to generate the new datasets was multiplied by seven different factors (Fs): 0.10, 0.15, 0.20, 0.5, 1.0, 2.0, and 5.0, with F = 1.0 being the predefined FiloGen configuration. This factor,
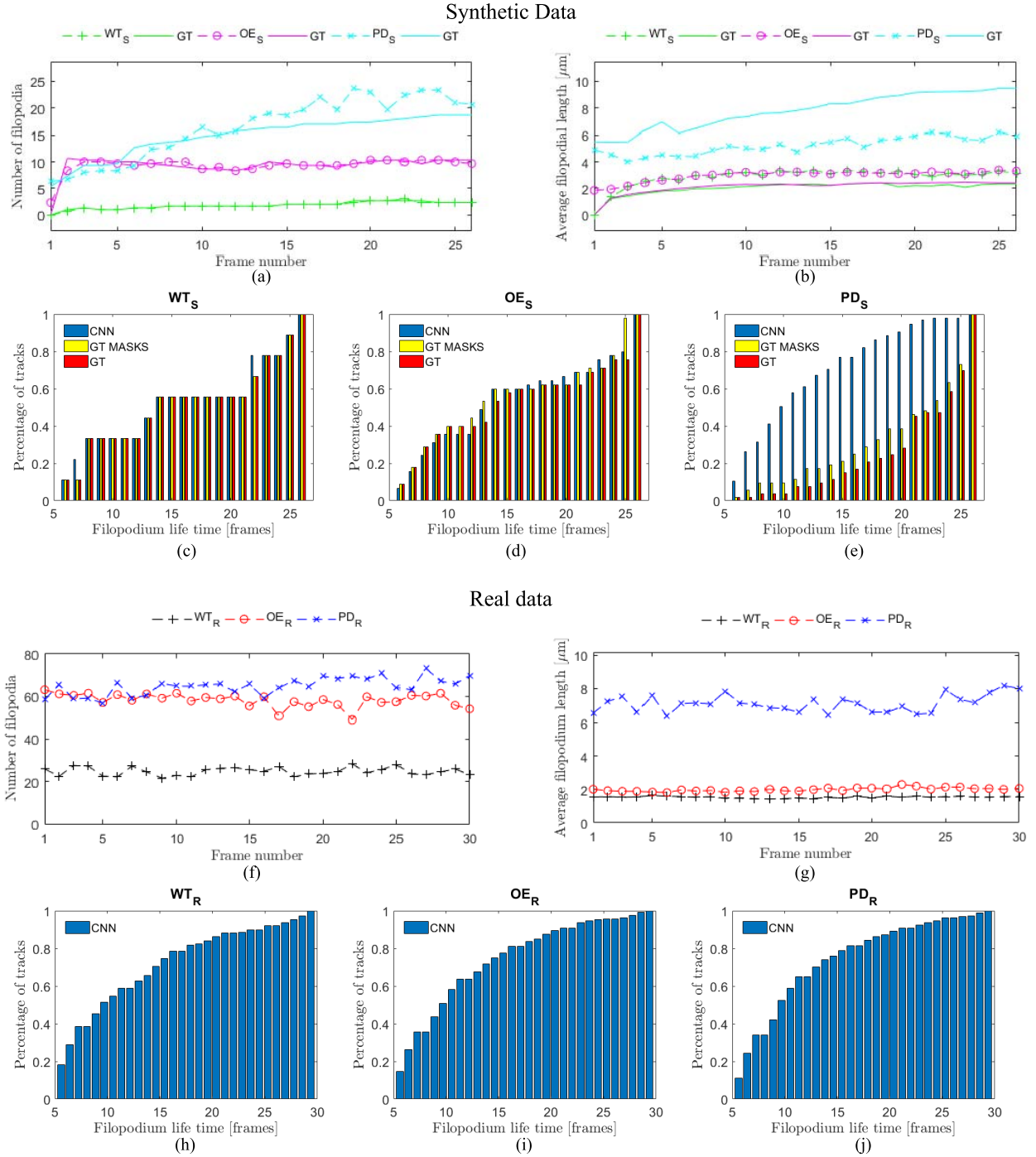
Fig. 6. Average temporal profiles of the number (a,f), average length (b,g), and lifetime cumulative histograms (c-e, h-j) of the filopodial extensions per frame for all WT_R, OE_R, PD_R, WT_S, OE_S and PD_S experiments respectively. The synthetic ground truth with continuous lines are shown for comparison (a-b). CNN method results shown.

along with the actual geometry of the cell, indirectly determines the SNR level of the generated image data. This approach is viable due to a predominance of photon shot (Poisson) noise in fluorescence microscopy and its dependence on the fluorescence level. All the new sequences were first generated using the PSF of a spinning disk confocal microscope equipped with a water Plan-Apochromatic 63x/1.20 used in the original datasets (Supplementary Figures 2 & 3). Also, from each cell geometry and for noise

factor F=1.0, three new sequences were generated using also the theoretical PSF (Huygens, SVI, the Netherlands) of a laser scanning confocal microscope equipped with a water Plan-Apochromatic 63x/1.40 objective lens, with the pinhole size fixed at three Airy units (Supplementary Figures 4 & 5).

We first compared the tracking results obtained on the new phenotypes with those obtained on the original synthetic phenotypes using the same multiplicative factor (F=1.00), i.e. with a similar average SNR (Supplementary Table V).

Our ANOVA analysis indicates that there are no significant differences between the detection results obtained on the new and the original phenotypes. Regarding the recall measure, the results are significantly better on the new phenotypes. We next evaluated the robustness of our workflow against increasing SNR levels (Supplementary Table VI). Our ANOVA analysis shows that the detection and recall measures are not significantly different above F=0.2. Finally, we compared the tracking results obtained for sequences generated with both PSFs, for a similar amount of noise (Supplementary Table VII). Our ANOVA analysis shows that for detection, the results obtained on the sequences generated with the PSF of a laser scanning confocal are better than those obtained on the generated using the PSF of the spinning disk confocal microscope. Supplementary Note 3 contains examples of correct and incorrect detections and tracking assignments.[7]

V. DISCUSSION

The pipeline presented and evaluated in this article shows a novel and effective approach to quantifying and tracking actin-rich filopodia protruded in 3D cancer cells. The contributions of our method to the state of the art are:

i. To the best of our knowledge, all previous attempts to quantify and track filopodial structures were based on cells lying on a 2D surface or on 2D maximum intensity projections of cells acquired in 3D. We instead, for the first time, analyze fully in 3D the morphology and dynamics of filopodial structures of cells embedded in a 3D matrix. This 3D setup is of much higher physiological significance than the 2D ones, as it has been shown that the mechanisms of cell migration significantly differ between 2D and 3D environments [27], [28]. Furthermore, by performing the analysis fully in 3D instead of in maximum intensity projections of the cells, we ensure a proper detection and quantification of filopodia with arbitrary orientation.

ii. We have developed and evaluated our segmentation and tracking pipeline using a unique dataset composed of both real and computer-generated cells with three distinct phenotypes in terms of filopodial morphology and dynamics. These datasets represent an objective benchmark for the evaluation of our method and are available to the scientific community for future algorithmic improvements.

iii. Our cell segmentation and filopodia detection pipeline makes use of *deep learning* methods, whose promising results have been recently highlighted in the area of particle [19] and cell tracking [18]. Indeed, we present the first use of CNNs for the segmentation of cells and filopodial tips in 3D. The complexity of properly training these networks, given the size and variability of the image data, cannot be overemphasized. In that regard, having access to synthetic data was instrumental to facilitating the correct training of the CNNs without the need for extensive, time consuming, highly subjective, manual annotations of the 3D+t image data.

---

[7]Supplementary materials are available in the supplementary files/multimedia tab

The presented CNN-based pipeline outperforms a state-of-the art method based on the minimization of the Chan-Vese model, previously shown as a very powerful segmentation tool for the segmentation of high-resolution 3D cells [21]. Indeed, the baseline CVS method belongs to the *tracking by contour* evolution cell tracking paradigm, which has been deemed especially appropriate for tracking of cells imaged at high resolution, provided that, as in this case, there is sufficient time resolution, and therefore cell overlap between consecutive video frames [18]. Furthermore, the CVS method takes full advantage of the 3D nature of the image data compared to our previous slice-by-slice segmentation method [16]. Here, we take advantage of the segmentation improvement provided by the CNN [17] to complete the task by addressing the detection and tracking of filopodial tips. In both tasks, a CNN trained indistinctly of the phenotype outperforms the results obtained using the CVS method optimized per phenotype using a genetic algorithm. On one hand, the CNN outperforms the CVS method in regard to the segmentation accuracy, both globally and for the real datasets, for which the weighting strategy of the CNN was specifically tuned. On the other hand, the CNN based pipeline globally outperforms the CVS based pipeline in terms of detection, recall and localization. The only exception is the detection of filopodial tips in the PD_S cells, where the CNN produces spurious detections. This is not unexpected since the weighing strategy of the CNN was optimized for the detection of filopodial tips in real cells. Regarding the quantification of the lifetime of the filopodia, the CNN provides accurate results for the WT_S and OE_S phenotypes. This is not the case for the PD_S phenotype, due to the above-mentioned detection problem. These results could be improved by specific CNN training and a more sophisticated linking algorithm.

We have also tested the robustness of our workflow on two intermediate synthetic phenotypes with different widths and lengths of filopodial protrusions that were not used in the training of the segmentation and tip detector CNNs. The results obtained on these new phenotypes are similar or better than those obtained on our original phenotypes, thus confirming the practical applicability of the method to previously unknown datasets. If such datasets to be tested were very different from the ones presented here and their results needed to be improved, both CNNs could be retrained with that new image data. We have also shown than the obtained tracking results are robust in a wide range of signal to noise ratios.

In summary, the CNN provides both quantitative (i.e., higher performance) and qualitative (i.e., generalizability, simplicity, self-learning capability) advantages compared to the baseline CVS method. Indeed, the trained CNN model can be used for the segmentation of cells with filopodia of different structural attributes, such as thickness, length, and level of branching, thus being directly applicable to natural scenarios when cells of multiple filopodial phenotypes co-exist in a tissue or culture. Contrarily, the optimization of the CVS algorithm reaches optimal values that are specific for each phenotype and are highly dependent of the uniformity of the morphology of the filopodia.

From the implementation point of view, we take advantage of an existing open-source platform [10], which provides a set of internal programming structures and a highly intuitive user interface. We have replaced or extended the analysis modules to 3D, to be able to use the interface to visualize the results provided by our method.

In future work, we intend to test our pipeline on new cell types and complex phenotypes. As shown, the presented CNN-based segmentation and tracking solutions provide a significant degree of generalization to somewhat similar phenotypes without the need for retraining the networks. Substantially different phenotypes might require training to achieve more accurate results. We also intend to improve the segmentation results in complex cases where the cortical staining does not correctly highlight the entire cell boundaries. Adding a second, cytoplasmic staining might help in that regard, although the improvement introduced would need to be evaluated also in the context of a possible negative effect on the acquisition time, thus temporal resolution, phototoxicity and experimental complexity. Finally, we intend to extend our pipeline to deal with more complex scenarios of multiple, mutually interacting cells with filopodial protrusions.

ACKNOWLEDGMENT

REFERENCES

[1] G. Jacquemet, H. Hamidi, and J. Ivaska, "Filopodia in cell adhesion, 3D migration and cancer cell invasion," *Current Opinion Cell Biol.*, vol. 31, no. 10, pp. 23–31, 2015.

[2] A. Haeger, K. Wolf, M. M. Zegers, and P. Friedl, "Collective cell migration: Guidance principles and hierarchies," *Trends Cell Biol.*, vol. 25, no. 9, pp. 556–566, 2015.

[3] K. H. Biswas and R. Zaidel-Bar, "Early events in the assembly of E-cadherin adhesions," *Exp. Cell Res.*, vol. 358, no. 1, pp. 14–19, 2017.

[4] P. K. Mattila and P. Lappalainen, "Filopodia: Molecular architecture and cellular functions," *Nature Rev. Mol. Cell Biol.*, vol. 9, no. 6, pp. 446–454, 2008.

[5] M.-F. Carlier, *Actin-based Motility: Cellular, Molecular and Physical Aspects*. Heidelberg, Germany: Springer-Verlag, 2010.

[6] T. Shibue, M. W. Brooks, M. F. Inan, F. Reinhardt, and R. A. Weinberg, "The outgrowth of micrometastases is enabled by the formation of filopodium-like protrusions," *Cancer Discovery*, vol. 2, no. 8, pp. 706–721, 2012.

[7] C. Ortiz-de-Solorzano, A. Muñoz-Barrutia, E. Meijering, and M. Kozubek, "Toward a morphodynamic model of the cell," *Signal Process. Mag.*, vol. 32, no. 1, pp. 20–29, 2015.

[8] V. Urbančič *et al.*, "Filopodyan: An open-source pipeline for the analysis of filopodia," *J. Cell Biol.*, vol. 216, no. 10, pp. 3405–3422, 2017.

[9] G. Jacquemet *et al.*, "FiloQuant reveals increased filopodia density during breast cancer progression," *J. Cell Biol.*, vol. 216, no. 10, pp. 3387–3403, 2017.

[10] D. Tsygankov, C. G. Bilancia, E. A. Vitriol, K. M. Hahn, M. Peifer, and T. C. Elston, "CellGeo: A computational platform for the analysis of shape changes in cells with complex geometries," *J. Cell Biol.*, vol. 204, no. 3, pp. 443–460, 2014.

[11] D. J. Barry, C. H. Durkin, J. V. Abella, and M. Way, "Open source software for quantification of cell migration, protrusions, and fluorescence intensities," *J. Cell Biol.*, vol. 209, no. 1, pp. 163–180, 2015.

[12] D. L. Coutu and T. Schroeder, "Probing cellular processes by long-term live imaging–historic problems and current solutions," *J. Cell Sci.*, vol. 126, no. 17, pp. 3805–3815, 2013.

[13] S. Constantino, C. B. Kent, A. G. Godin, T. E. Kennedy, P. W. Wiseman, and A. E. Fournier, "Semi-automated quantification of filopodial dynamics," *J. Neurosci. Methods*, vol. 171, no. 1, pp. 165–173, 2008.

[14] V. Tsimashchuk, W. J. Godinez, M. Lehmann, and K. Rohr, "Analysis of viral surfing based on fluorescence microscopy imaging and automatic tracking," in *Proc. IEEE Int. Symp. Biomed. Imag.*, May 2012, pp. 378–381.

[15] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 32–45, May 2006.

[16] M. Maška, X. Morales, A. Muñoz-Barrutia, A. Rouzaut, and C. Ortiz-de-Solórzano, "Automatic quantification of filopodia-based cell migration," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 668–671.

[17] C. Castilla, M. Maška, D. V. Sorokin, E. Meijering, and C. Ortiz-de-Solórzano, "Segmentation of actin-stained 3D fluorescent cells with filopodial protrusions using convolutional neural networks," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 413–417.

[18] V. Ulman, M. Maška, and C. Ortiz-de-Solórzano, "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, pp. 1141–1152, Oct. 2017.

[19] Y. Yao, I. Smal, and E. Meijering, "Deep neural networks for data association in particle tracking," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 458–461.

[20] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, and T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput., Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2016, pp. 424–432.

[21] M. Maška, O. Daněk, S. Garasa, A. Rouzaut, A. Muñoz-Barrutia, and C. Ortiz-de-Solórzano, "Segmentation and shape tracking of whole fluorescent cells based on the Chan–Vese model," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 995–1006, Jun. 2013.

[22] D. V. Sorokin, I. Peterlík, V. Ulman, D. Svoboda, and M. Maška, "Model-based generation of synthetic 3D time-lapse sequences of motile cells with growing filopodia," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 822–826.

[23] D. V. Sorokin *et al.*, "FiloGen: A model-based generator of synthetic 3D time-lapse sequences of single motile cells with growing and branching filopodia," *IEEE Trans. Med. Imag.*, to be published, doi: 10.1109/TMI.2018.2845884.

[24] M. Maška and P. Matula, "Particle tracking accuracy measurement based on comparison of linear oriented forests," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Oct. 2017, pp. 11–17.

[25] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1007–1019, Aug. 2004.

[26] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.

[27] M. K. Driscoll and G. Danuser, "Quantifying modes of 3D cell migration," *Trends Cell Biol.*, vol. 25, no. 12, pp. 749–759, 2015.

[28] P. Friedl and K. Wolf, "Plasticity of cell migration: A multiscale tuning model," *J. Cell Biol.*, vol. 188, no. 1, pp. 11–19, 2010.