# SEGMENTATION AND MODELLING OF HELA NUCLEAR ENVELOPE

*C. Karabağ⋆, M.L. Jones†, C.J. Peddie†, A.E. Weston†, L.M. Collinson† and C.C. Reyes-Aldasoro⋆*

⋆School of Mathematics, Computer Science and Engineering, City, University of London, UK
†Electron Microscopy Science Technology Platform, The Francis Crick Institute, UK

## ABSTRACT

This paper describes an algorithm to segment the 3D nuclear envelope of HeLa cancer cells from electron microscopy images and model the volumetric shape of the nuclear envelope against an ellipsoid. The algorithm was trained on a single cell and then tested in six separate cells. To assess the algorithm, Jaccard similarity index and Hausdorff distance against a manually-delineated gold standard were calculated on two cells. The mean Jaccard value and Hausdorff distance that the segmentation achieved for central slices were $98\%$ and $4$ pixels for the first cell and $94\%$ and $13$ pixels for the second cell and outperformed segmentation with active contours. The modelling projects a 3D to a 2D surface that summarises the complexity of the shape in an intuitive result. Measurements extracted from the modelled surface may be useful to correlate shape with biological characteristics. The algorithm is unsupervised, fully automatic, fast and processes one image in less than 10 seconds. Code and data are freely available at https://github.com/reyesaldasoro/Hela-Cell-Segmentation and http://dx.doi.org/10.6019/EMPIAR-10094.

***Index Terms*—** Segmentation, HeLa, Active contours.

## 1. INTRODUCTION

Cell segmentation and classification has been an important problem for many years [1] and has attracted considerable attention in clinical practice and research, as the presence or absence of cells or their characteristics like size or shape, could be important indicators for presence or severity of a disease [2]. Yet, despite considerable progress, automatic cell segmentation remains challenging and despite the significant disadvantages of time and inter- and intra-user variability, manual segmentation remains widely used [3].

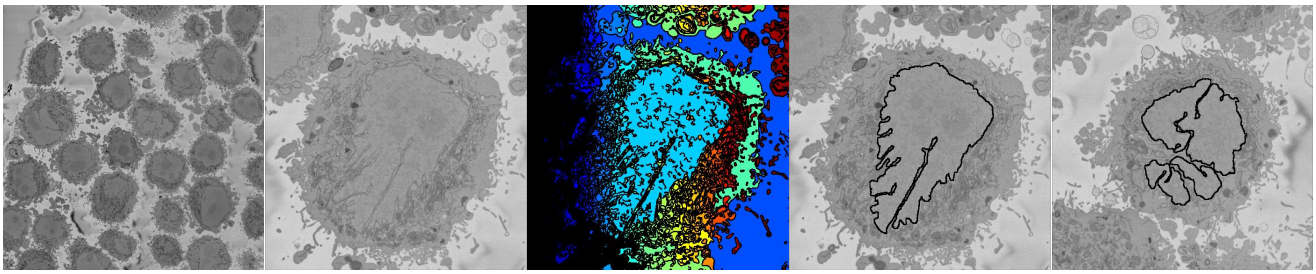
Segmentation of cells and particularly of nuclei and nuclear envelope (NE), which is the structure of interest of this paper, is highly dependent on the contrast, signal to noise ratio, complex morphological structures, and the resolution of the imaging. Thus, techniques that work at lower resolutions (such as those use for immunohistochemistry), like watersheds, are not immediately applicable at higher resolutions such as those provided by electron microscopy (EM). Furthermore, structures observed with EM display far more complex morphological structures, and many times with lower contrast than those observed in light and fluorescence microscopy. Serial section EM provides contiguous images of a sample [4] that correspond to the volumetric sample, thus the data is 3D and complicates both storing and transferring data as well as processing and interpreting it [5]. Deep learning methodologies [6] have become a popular tool for segmentation and classification, however, they require very large training data sets [7, 8], as well as significant computational power to train them, which may not always be available. Another segmentation alternative is called *citizen science* where an *army of non-experts* [9] are recruited to provide human techniques of segmentation or classification through web-based interfaces. Whilst these results are valuable, they take considerable time as they depend on volunteers (or paid workers like Amazon's Mechanical Turk) and they do not guarantee a correct answer.

In this work, an automatic, unsupervised algorithm to segment the NE of HeLa cells is described. The algorithm presented exploits intensity variations of the structures of HeLa cells following [10]. Preliminary work segmented only one cell and, as an extension, this work now segments 7 different cells with disconnected regions which is the main contribution and some metrics were extracted from the surfaces by modelling the 3D surface of the NE against a 3D ellipsoid.

## 2. MATERIALS AND METHODS

### 2.1. HeLa cells preparation and acquisition

Details of the acquisition have been published previously [10], but briefly, the data sets consisted of EM images of cancerous HeLa cells. The method of the National Centre for Microscopy and Imaging Research (NCMIR)[11] was followed and wild type HeLa cells were prepared and embedded in Durcupan resin. Serial blockface scanning electron microscopy (SBF SEM) was used to obtain $518$ EM images of $8192 \times 8192$ pixels each. Voxels size was with $10 \times 10 \times 50$ nm with intensity $[0 - 255]$ (Fig. 1a). The data was acquired at higher bit-depth (32 bit or 16 bit) and reduced to 8 bit after contrast/histogram adjustment. Seven individual cells were manually cropped as volumes of interest of 300 slices with dimensions $(n_h, n_w, n_d) = (2000, 2000, 1)$ each (Fig. 1b) and saved as single channel *TIFF* files.



(a) (b) (c) (d) (e)

**Fig. 1**. (a) One representative slice of a 3D stack acquired with an EM containing numerous HeLa cells. (b) Region of Interest (ROI) with one HeLa cell centred. The nucleus is the large and fairly uniform region in the centre and it is surrounded by the nuclear envelope which is darker than the nucleus. (c) Superpixels obtained with the algorithm. (d) Final segmentation of the NE shown by the thick black line. (e) Segmentation of a different slice showing several disjoint nuclear regions.

### 2.2. Ground truth (GT)

The nuclear envelope of two cells were segmented manually (Figs. 2a,d) by delineating the NE slice-by-slice. The segmentations were performed separately by different persons without knowledge of each other or the algorithm. Disjoint regions of the nucleus were assessed by scrolling up and down slices to determine if the regions were part of the nucleus and connected at upper or lower slices.

### 2.3. Automatic segmentation of the nuclear envelope

In this work, the NE of seven different cells were analyzed. The algorithm was developed and trained on one cell, for which GT existed, (Fig. 2a) and tested on the NEs of the other six cells (Figs. 2d, 3a-e) of which only one had a corresponding GT (Fig 2d). All processing and visualisation was performed in MATLAB® (The Mathworks™, Natick, USA) and the code is available open-source. Images were low-pass filtered with a Gaussian kernel with size $h = 7$ and standard deviation $\sigma = 2$ to remove high frequency noise, which gave a grainy appearance to the images. The framework exploited the darker intensity of the NE as compared with its neighbouring cytoplasm and nucleoplasm by Canny edge detection [12] to detect abrupt changes of intensity. The edges were dilated to connect disjoint edges, which were part of the NE, due to intensity variations of the envelope itself. The connected pixels not covered by the dilated edges were labelled to create a series of superpixels (Fig. 1c). The superpixel size was not restricted as large superpixels covered the background and nucleus itself. Morphological operators were used to: remove regions in contact with the edges of the image, remove small regions and fill holes inside larger regions. The central superpixel was selected as the nucleus and further morphological operators were applied to close the jagged edges (Figs. 1d, 2a,b,d). To improve the 2D-based methodology presented in [10], the algorithm considered the data as 3D, so that each slice was processed with the knowledge of the segmentation of its upper/lower neighbouring slice. The central slice, which was assumed to be the one in which the nucleus would be largest, was processed first, and that segmentation was propagated up and down in the same way that a human would scroll up and down to decide if a disjoint nuclear region was connected above or below the current slice of analysis. By using neighbouring segmentations as input parameters to the segmentation, the algorithm was able to identify disjoint nuclear regions as a single nucleus (Fig. 1e). Finally, the NE was obtained as the boundary of the nucleus. Optimal parameters of the algorithm were obtained through sensitivity analyses (results not shown).

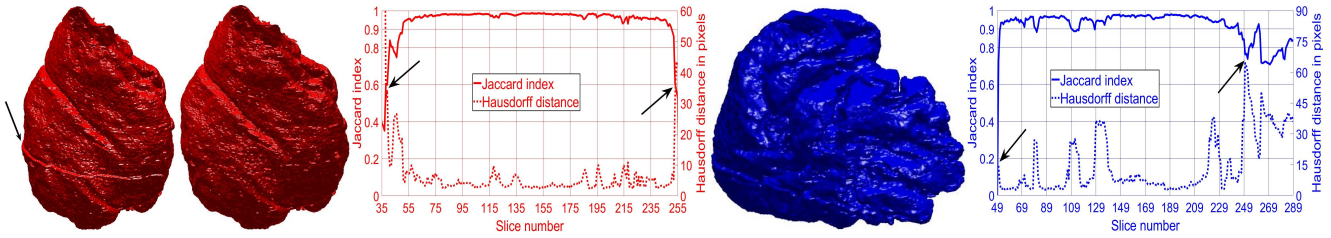
### 2.4. Quantitative comparisons

The segmentation results were compared against the GT with the Jaccard similarity index (JI) of intersection over union of nuclear area [13] and the Hausdorff distance [14], the maximum of the set of shortest distances between corresponding points the nuclear envelope. To compute JI, the manual delineation in each slice was morphologically closed to generate a region rather than a line. The Hausdorff distance between the GT and NE were calculated for each slice.

### 2.5. Active contours

For comparison, the nuclear envelope was segmented with Chan-Vese active contours method [15]. The function changes its parameters based on one of three states: Shrink, Grow, or Normal. One of the three states and its parameters were chosen empirically through numerous tests. The active contour was run once, with a set of parameters, then the Jaccard index was calculated. The parameters were adjusted (contraction bias=1.5, smooth factor=-0.4, iterations=5000) and the active contour was run again to obtain foreground and background.

The highest JI value obtained with active contours was $75\%$ at 2200 iterations, which was lower than those obtained with the algorithm and presented below. The other two states of active contours, Shrink and Normal, were also implemented but provided worse results than Grow.




(a) (b) (c) (d) (e)

**Fig. 2**. (a,b,d) Volumetric segmentation of two nuclear envelopes: training cell (a,b) and testing cell (d). The arrow in (a) indicates an artifact due to a slight shift in the acquisition, which was corrected with a pre-alignment and not present in (b). (c,e) Jaccard similarity index (solid line) and Hausdorff distance (dashed line) for first and second cell respectively. Notice that when Jaccard index decreases Hausdorff distance increases (arrows) as the smaller the Hausdorff distance between two shapes the greater is their degree of resemblance.

### 2.6. Nuclear envelope shape modelling

The shape of the segmented NE was modelled against a 3D ellipsoid (Fig. 4a). The ellipsoid was adjusted to have the same volume as the nucleus. The surfaces of the ellipsoid and the nucleus were subsequently compared by tracing rays (Fig. 4b) from the centre of the ellipsoid and the distance between the surfaces for each ray was calculated (Fig. 4c). It was assumed that when the nucleus surface was further away from the centre, the difference was positive.

## 3. RESULTS

The algorithm was designed and trained on one cell from which the parameters were derived. The NE of seven cells were then segmented with the algorithm Figs. 2a,b,d, Fig. 3) with accurate results for the two sets for which GT was available and good visual assessment of the remaining 5. The shapes of the final segmentations show the complexity of the nuclear envelopes with rather convoluted notches and invaginations. It is speculated that these shapes may have biological significance.

Whilst segmenting the first cell, a displacement artifact, which is assumed was caused by an external vibration to the microscope, was detected (arrow in Fig. 2a). The displacement was corrected (Fig. 2b) and since it was present on both images and GT had no impact on the JI or Hausdorff distance.

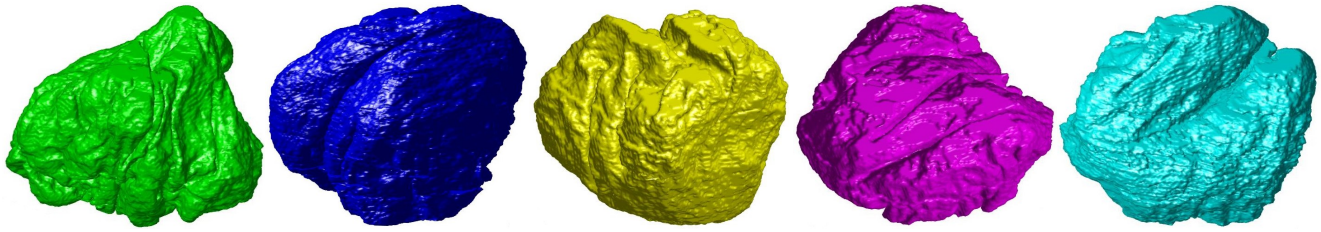
Comparison with manual segmented ground truth reported mean Jaccard Similarity Index (JI) $98\%$ with mean Hausdorff distance $4$ pixels for the first cell and $94\%$ and $13$ pixels for the second on the central slices. JI decreased and Hausdorff distance increased towards the top and bottom of the cells as the structure was considerably more complex and the areas become much smaller (Figs. 2c,e). Active contours only on one slice (slice 118/300) of the first cell reported a maximum JI of 75%, with 2200 iterations and required 27 minutes. On the other hand the algorithm described segmented each slice in $\sim$8 seconds.

The comparison between the model ellipsoid and the whole nucleus reported a JI of $71\%$ (Fig. 4a). This value indicates a relative departure from a sphere and it is speculated that the JI could be related with biological characteristics of cells. In addition, the measurements of distance from the nucleus to the ellipsoid showed rougher and smoother regions (Figs. 4b,c, d). The surface corresponding to the distance from the nuclear envelope to a model ellipsoid (Fig. 4d) showed graphically the hollow and prominent regions of the cell, but more important, elements such as a notch (solid red arrow) or ruggedness (dashed green arrow) can be an indication of NE break down or remodelling. An advantage of this modelling is that visually it is easier to assess a single 2D image than a 3D volumetric surface.

## 4. DISCUSSION

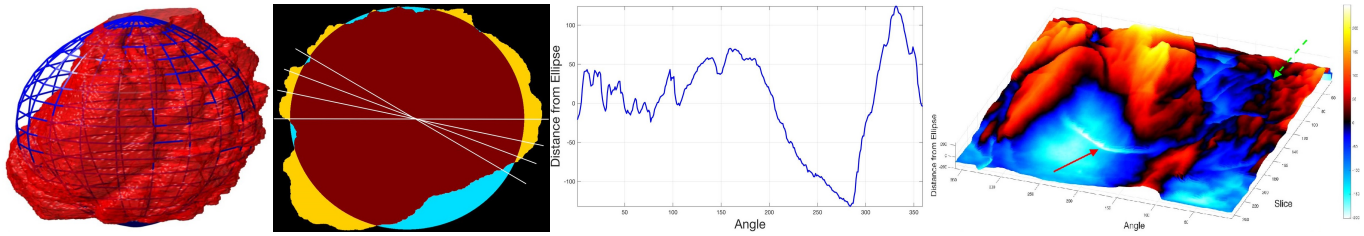
In this work an automated algorithm to segment the nuclear envelope of HeLa cell and the modelling of the volumetric shape against an ellipsoid were proposed. Seven cells were successfully segmented. The algorithm is fully automated, unsupervised, segmented each slice in approximately 8 seconds and does not require training data. Two different similarity metrics, Jaccard Index and Hausdorf distance, were computed to test the algorithm against the GT and promising segmentation results were obtained as indicated by these metrics. Active contours method was also implemented and the results were inferior to the algorithm.

The main contributions of this work are: (a) a framework to segment the nuclear envelope of HeLa cells from SBF SEM images, which includes disjoint regions of the nucleus. (b) The modelling of the NE surface against an ellipsoid, which could reveal interesting biological characteristics of the nucleus. (c) The 2D maps of the NE surface, which can provide an easier way to assess the characteristics of a 3D structure. The framework is not restricted of course to HeLa cells. Other cells from any electron imaging modality could be also analysed. Future work will consider the analysis of the whole



(a) (b) (c) (d) (e)

**Fig. 3**. Volumetric surfaces of five different nuclear envelopes. Notice the variation of each NE with different rugosity, notches and invaginations.




(a) (b) (c) (d)

**Fig. 4**. (a) Rendering of the nuclear envelope (red) against the model ellipsoid (blue mesh). (b) llustration of distance calculations by ray tracing in one slice. Yellow regions correspond to the nucleus outside the ellipsoid, cyan regions where nucleus inside the ellipsoid. (c) Measurements obtained along the boundary. (d) Surface corresponding to the distance from the nuclear envelope to a model ellipsoid. Solid red arrow indicates a notch, dashed green arrow shows rugged region.

cell, other cellular structures and the chromatin in the nucleus. Furthermore, other methodologies such as Deep learning will be explored once the sufficient training data is available.

## 5. REFERENCES


[1] Z. Wang et al., "Generalizing cell segmentation and quantification," *BMC Bio*, vol. 18, no. 1, pp. 189, 2017.

[2] M.N. Gurcan et al., "Histopathological Image Analysis: A Review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[3] P. Bajcsy et al., "Survey statistics of automated segmentations applied to optical imaging of mammalian cells," *BMC Bioinformatics*, vol. 16, no. 330, 2015.

[4] L. Ostroff et al., "Electron Microscopy at Scale," *Cell*, vol. 162, pp. 474–475, July 2015.

[5] C.J. Peddie et al., "Exploring the third dimension: Volume electron microscopy comes of age," *Micron*, vol. 61, pp. 919, Jun 2014.

[6] Y. LeCun et al, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[7] J. C. Caicedo et al., "Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images," *IEEE Rev. Bio. Eng*, vol. 2, pp. 147–171, 2018.

[8] T. M. Quan et al., "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," *arXiv:1612.05360*, 2016.

[9] J.L. Schnoor, "Citizen science," *Environmental Science & Tech.*, vol. 41, no. 17, pp. 5923–5923, Sep 2007.

[10] C. Karabağ et al., "Automated Segmentation of HeLa Nuclear Envelope from Electron Microscopy Images," in *MIUA*, July 2018, vol. 894, pp. 803–806.

[11] T. J. Deerinck et al., "NCMIR: A new protocol for preparation of biological specimens for serial block-face SEM microscopy," 2010, Accessed: 2018-10-08.

[12] J. Canny, "A Comp. Approach to Edge Detection," *IEEE Trans PAMI*, vol. 8, no. 6, pp. 679–698, 1986.

[13] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bull. Soc. Vaudoise Sc. Nat.*, vol. 37, pp. 547–579, 1901.

[14] I.S. Kim et al., "Computing the hausdorff distance between two sets of parametric curves," *Commun. Korean Math. Soc.*, vol. 28, no. 4, pp. 833–850, 2013.

[15] T.F. Chan et al., "Active contours without edges," *IEEE Trans Imag Proc*, vol. 10, no. 2, pp. 266–277, Jan 2001.