On Benchmarking Cell Nuclei Segmentation Algorithms for Fluorescence Microscopy

Frederike Wirth, Eva-Maria Brinkmann and Klaus Brinker

Hamm-Lippstadt University of Applied Sciences, Marker Allee 76-78, 59063 Hamm, Germany

Keywords: Medical Image Analysis, Computer Vision, Fluorescence Microscopy, Cell Nuclei Segmentation, Segmentation Benchmarking, Watershed Segmentation.

Abstract: Cell nuclei detection is an essential step in the context of many image analysis tasks related to microscopy images and therefore also plays a role in highly topical fields of research like the development of personalised immunotherapy approaches against several types of cancer. Motivated by this observation, a whole zoo of advanced methods to accomplish cell nuclei segmentation has been proposed in recent years. This development in turn stresses the need to set up a well-justified and reproducible standard of comparison for the evaluation of these sophisticated approaches. In this paper, we describe how such a reference framework based on standard seeded watershed segmentation for fully automatic cell nuclei detection can be obtained. In particular, we provide a detailed review of a publicly available dataset, give a detailed account of the methods and evaluation measures we consider to enable the highest possible reproducibility of our results and discuss the suitability of different variants of seeded watershed segmentation for the mentioned purposes.

1 INTRODUCTION

In 2018, the Nobel Prize in Physiology or Medicine was awarded to the two immunologists J. P. Allison and T. Honjo crediting their seminal biochemical discoveries that laid the foundation for new approaches to cancer therapy using so-called immune checkpoint inhibitors (ICIs). In recent years, this novel type of treatment led to groundbreaking advances in the fight against cancer, where the general idea underlying any immunotherapy approach is to activate the patient's immune system to attack tumour cells and ideally kill the cancer (cf. e.g. (Kelly, 2018)). While first long term studies (cf. e.g. (Schadendorf et al., 2015; Antonia et al., 2017)) corroborate that for various types of cancer an immune checkpoint inhibitor-based therapy may lead to a significantly longer progression-free survival (PFS) and higher 5-year overall survival (OS) rates, it is still an open research question to predict which patients will actually benefit from this novel cancer immunotherapy or might encounter severe side effects.

In order to make a contribution to a more personalised medical treatment in this area, we plan to build a predictive model based on fluorescence microscopy imaging using multiple biomarkers. Within the overall processing pipeline, the first step is to detect the cell nuclei. Hence, the aim of this paper is to set up a reference framework for cell nuclei segmentation.

Relying on well-established standard methods, variants of watershed segmentation, to be precise, that we apply to publicly available cell image data, we create a reliable benchmark for the evaluation of more sophisticated segmentation algorithms. These constitute the starting point to create a personalised treatment plan.

The remainder of this paper is organised as follows: First, we describe the two datasets that we consider throughout this work and explain how this data can be converted into a ground truth segmentation for later evaluation. Subsequently, we recall the general idea of the watershed segmentation method and give a detailed account about our implementation of several variants of this well-established approach. Even more, we recapitulate the definition of common performance measures. Based on these measures as well as a visual inspection of the segmentation results, we eventually compare the previously described variants of watershed segmentation, thus providing a reproducible evaluation baseline for more involved segmentation approaches.

2 THE DATASET

In the following, we will start with a description of the used dataset including basic information collected in a manual review of the data. Afterwards, we will

explain how to convert the images into a more appropriate representation. All cell lines and a hand-labelled image collection were published by the group of R. F. Murphy[1] in order to facilitate further research (cf. also (Coelho et al., 2009).

2.1 Data Review

The dataset that is used in this paper contains two cell lines, U20S and NIH3T3. The U20S cell line was originally created and used by Peng et al. for the development of a pattern unmixing algorithm (Peng et al., 2010). The images contained in this cell line show human osteosarcoma cells dyed with Hoechst 33342 (Peng et al., 2010). Recalling that in our previously sketched overall research project we are interested in the analysis of fluorescence microscopy images of human cancer cells, the images of the U20S cell line seem particularly appropriate as a preparation for further work in the context of this project. The origin of the NIH3T3 cell images, is a collection using the method of Osuna et al. (Osuna et al., 2007). These mouse embryo cells where cultured by Todaro and Green (Todaro, 1963) and also dyed with the fluorescence stain Hoechst 33342 (Osuna et al., 2007). Looking at the U20S cell line, we notice that some of the cells are clustered, which renders the segmentation of single cell nuclei particularly challenging. The number of cells per image varies between 14 and 43. Some of the images contain bright artefact spots as part of the background. The mean brightness of the cells shown in the images of the NIH3T3 cell line is lower than in the first image collection. In addition, the former contains even brighter artefacts than the U20S images. We also note that the cells of the NIH3T3 line are less clustered and the number of cells varies between 18 and 56.

2.2 Data Conversion

Beside the fluorescence microscopy nuclei images (50 per cell line), the data also contains five images per cell line that are hand segmented by A. Shariff as well as 48 images and 49 images, respectively, that are hand segmented by L. P. Coelho. The three missing images were rejected by L. P. Coelho as being not in-focus. The hand segmented masks were provided as the original microscopy image with the cell contours marked in red. However, to compute further evaluation measurements, it is more appropriate to represent the images in a manner such that the background is labelled as zero and every cell has

a unique integer label. Therefore, a flood fill algorithm was used to convert the cell boundary images. Care must be taken because some of the borders do not completely surround the cells, so the flood fill algorithm might consider these cells to be part of the background. This problem can be solved by an accurate review of the data, where existing holes in the cell boundaries are manually closed. Another difficulty of the automatic image conversion are tiny artefacts with a size of one pixel that might be counted as a single cell. To avoid these small cells to distort the results, they can also be manually deleted or removed by an automatic small object filter, as it is e.g. provided by *scikit-image*[2]. It is important to note that there are differences between the reference images manually segmented by L. P. Coelho and the ones segmented by A. Shariff. Those discrepancies are i.a. caused by the way cells touching the border of the image are handled. As a consequence, we decided to delete every border cell to avoid any ambiguities.

3 SEGMENTATION METHODS

As mentioned before, the goal of this paper is to describe the creation of an evaluation baseline for the future assessment of state-of-the-art segmentation approaches with a particular eye on cell nuclei detection in fluorescence microscopy images.

In the context of benchmarking, a key point is the easy reproducibility of results. Hence, we decided to only draw on standard methods, for which fast and reliable implementations in widely-accessible computer vision libraries such as *OpenCV* or *scikit-image* exist. A well-established class of approaches that not only satisfy this criterion but, in addition, are known to be able to separate touching and overlapping objects (cf. (Russ and Neal, 2015, p. 483) and references therein) rendering them particularly suited for the detection of clustered cells in microscopy images are segmentation algorithms that are based on the so-called *watershed transformation*.

3.1 Watershed Transformation

An important first step for understanding the watershed transformation in image processing is a re-interpretation of a given single-channel image as a topological relief with the intensity values being regarded as physical elevation. Then, the central idea of the watershed transformation first introduced as a morphological tool in (Digabel and Lantuéjoul, 1978)

---

[1]http://murphylab.web.cmu.edu/data

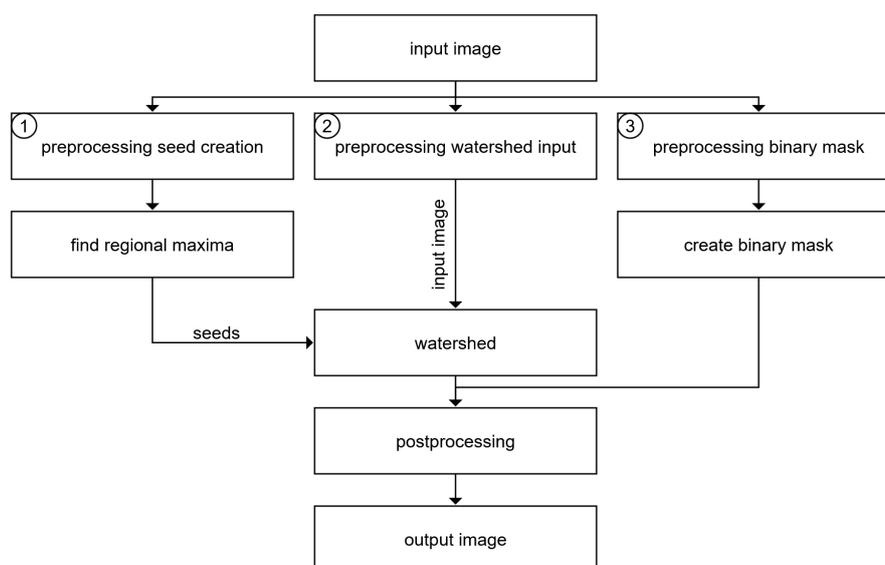[2]https://scikit-image.org/ (version 0.15.0)

Figure 1: Schematic representation of segmentation pipeline based on the watershed transformation.

and made applicable to digital images by Vincent and Soille (Vincent and Soille, 1991) can be regarded as the straightforward adaptation of the concept of watersheds in geography to this setting. *Watersheds* are the lines that separate adjacent so-called *catchment basins*, the latter describing the area of land from which water flows to a deeper water reservoir like a river, a lake or an ocean.

Given the above identification of image and topological relief, the characterization of catchment basins and watersheds is a means to assign every pixel in the image either to a specific region or to a line that separates two regions. In order to determine the catchment basins, we now imagine that the relief is gradually filled with water. More precisely, we assume that starting from the lowest points of the relief, i.e. the smallest intensity values of the image, the water pours out through local minima, which in this context are commonly referred to as *seeds.* By this gradual immersion of the relief more and more catchment basins around the seeds emerge and grow, this way iteratively assigning adjacent pixels to regions. Whenever the catchment basins flooded by two different seeds are about to fuse, the contact pixels are labelled as watershed lines such that we finally obtain a full decomposition of the image into various regions of cohesive pixels separated by closed contours. As a consequence, the watershed transformation is often applied in the context of image segmentation.

3.2 Implementation

In order to create a baseline for new cell segmentation algorithms, we implemented a seeded watershed pipeline that is based on methods of the following standard python libraries: *NumPy*[3], *OpenCV*[4], *scikit-image*[5]. Figure 1 illustrates the process of this segmentation pipeline that starts with an optional preprocessing of the input image before the seed detection takes place, which is based on the computation of regional maxima. The latter is defined as a pixel with an intensity value v that is not directly located next to a pixel with an intensity value higher than v. Additionally, it is not connected to such a pixel through chaining pixels with an intensity value equal to v. The input image of the watershed method might also be preprocessed either with a Gaussian blur or with a distance function as proposed in a more recent online tutorial of L. P. Coelho.[6] The watershed transformation is not able to distinguish between cells and the image background. As a consequence, it is necessary to subtract a mask created with the mean threshold afterwards. The mask is not generated with the original image but with a blurred version of it. Finally, we applied some postprocessing steps to this mask. By analogy with the conversion of the ground truth data, we deleted all

---

[3]https://numpy.org (version 1.16.5)

[4]https://opencv.org (version 3.4.2)

[5]https://scikit-image.org/ (version 0.15.0)

[6]https://github.com/luispedro/python-image-tutorial/blob/master/Segmenting%20cell%20images%2020(fluorescent%20microscopy).ipynb

Table 1: Overview of the various implemented variants for watershed segmentation considered in this paper (cf. also Figure 1).

variant of watershed segmentation	① σ-value for seed computation	② watershed input image	③ threshold for binary mask creation
variant 1	12	original intensity image	12
variant 2	12	blurred intensity image ($\sigma = 6$)	12
variant 3	19	original intensity image	12
variant 4	19	original intensity image	4
variant 5	19	threshold with t_{mean} of blurred intensity image ($\sigma = 2$), then distance transform	4

boundary cells from the segmentation masks. Moreover, we applied an object filter that removes all objects smaller than the smallest cell of the respective cell line in the reference dataset.

Looking at the full segmentation pipeline (see Figure 1), there are three possible points of adjustment: the value of σ controlling the amount of Gaussian blur in the image used for computing the seeds (see number 1 in Figure 1), the preprocessing steps of the watershed input image (see number 2 in Figure 1) and also the σ-value of the Gaussian blur for the creation of the threshold mask (see number 3 in Figure 1).

We consider the following segmentation pipelines as summarised in Table 1: Variant 1 is an implementation of the settings proposed by Coelho et al. (Coelho et al., 2009). The watershed transformation of variant 2 works on a blurred version of the cell images. In variant 3, the σ-value that controls the blurring of the original input image before the maxima detection is increased, which lowers the absolute number of watershed seeds. Variant 4 is based on the same generation of seeds but the σ-value used for the creation of the binary mask is decreased. The watershed input image of variant 3 and 4 is the original image. For variant 5 the input of the watershed image is a distance transform of the thresholded binary image, where each pixel is assigned the shortest (Euclidean) distance to the background as labelled by the binary mask.

4 EVALUATION

Regarding the manual segmentations by L. P. Coelho as ground truth, we can compare the performance of the various variants of watershed segmentation explained above by employing several well-known supervised evaluation methods. These approaches can

be categorised in pixel-based and object-based methods. The *Rand Index (RI)* (Rand, 1971) and the *Jaccard Index (JI)* (Jaccard, 1901) both belong to the former class. They compare the assignment of every pixel in the segmented and reference image. Their disadvantage is that they take no account of information about the location of the cell border. That is why we additionally calculate the *Hausdorff Distance* as described by Chalana and Kim (Chalana and Kim, 1997, and references therein)) as well as the *Normalised Sum of Distances (NSD)* proposed by Coelho et al. (Coelho et al., 2009). A third group of evaluation measures are the four error counting values *split*, *merged*, *added* and *missing*. They refer to the mistakes in the cell images that are intuitively seen by a human observer and have been used in similar works as e.g. (Baltissen et al., 2018; Coelho et al., 2009). Here and in the following, *R* denotes the reference image, i.e. the ground truth, and *S* refers to the associated segmented image created with one of the mentioned algorithms.

4.1 Rand Index and Jaccard Index

The Rand Index (Rand, 1971) and the Jaccard Index (Jaccard, 1901) were developed for the evaluation of clusterings. In a certain sense, they both measure to what extend the reference image and the segmented image agree about pixels being assigned to the same object or different objects. Let S_i be the intensity value of a pixel in the segmented image S and S_j be the intensity value of a second pixel in S. Analogously, R_i and R_j are the intensity values of a pixel pair in the reference image with the same coordinates as S_i and S_j. We consider every possible combination of two different pixels, disregarding order, i.e. every combination of S_i and S_j with $i \neq j$ in the segmented image. All these possible pixel pairs are then categorised in four groups, labelled as A, B, C and D:

- A: $S_i = S_j$ and $R_i = R_j$,

- B: $S_i \neq S_j$ and $R_i = R_j$,

- C: $S_i = S_j$ and $R_i \neq R_j$,

- D: $S_i \neq S_j$ and $R_i \neq R_j$.

In other words, the categories *A* and *D* include every pixel pair, where the two segmentations agree about assigning both pixels to the same object or not. On the contrary, *B* and *C* contain the pixel pairs where the reference and the segmented image disagree about the pixels at position *i* and *j* being part of the same object. Moreover, we let *a*, *b*, *c* and *d* denote the number of pixel pairs in the categories *A*, *B*, *C* and *D*, respectively.

Given these notations, the *Rand Index (RI)* is defined as the quotient

$$\text{RI}(R,S) = \frac{a+d}{a+b+c+d} = \frac{a+d}{\binom{n}{2}}, \qquad (1)$$

where *n* denotes the total number of pixels either in the segmented image *S* or the reference image *R*. A Rand Index of 1.0 implies a complete overlap of the objects in *R* and *S*, while a small overlap between the segmented and the reference regions corresponds to a lower Rand Index.

The *Jaccard Index (JI)* given by

$$\text{JI}(R,S) = \frac{a+d}{b+c+d} \qquad (2)$$

is based on the same classification of pixel pairs as explained above. The upper limit of the Jaccard Index depends on the number and size of cells in comparison to the image size. Consequently, it is only appropriate to compare Jaccard Indices taken from results created with identical image data.

4.2 Hausdorff Distance and Normalised Sum of Distances

The pixel-based metrics have a serious disadvantage: they are quite sensitive to small variations of the border position. Other well-known segmentation metrics deal with this problem by using the distance between the reference and the segmented borders as a measure. The *Hausdorff Distance (HD)* proposed by Bamford (Bamford, 2003), is defined as the maximum of the set of shortest distances between two shapes, or as described by Coelho et al. (Coelho et al., 2009)

$$\text{HD}(R,S) = \max\{D(i) : S_i \neq R_i\}, \qquad (3)$$

where $D(i)$ is the distance of every pixel to the border of the reference object. The Hausdorff Distance is calculated for every segmented object and its assigned

reference object. To aggregate the results for the entire image, the average Hausdorff Distance of all objects is calculated. A similar approach is the *Normalised Sum of Distances (NSD)* which is described as

$$\text{NSD}(R,S) = \frac{\sum_i [\![R_i \neq S_i]\!] * D(i)}{\sum_i D(i)}, \qquad (4)$$

where $D(i)$ is again the distance of every pixel to the border of the reference object (Coelho et al., 2009). In the same manner as described above for the Hausdorff Distance, we then calculated the average NSD value of all segmented objects.

4.3 Error Counting Metrices

The error counting metrics *split*, *merged*, *added* and *missing* belong to the class of object-based measurements and depend on the detection of objects in the two images. As a preparation step, every object of *S* is assigned to an object in *R*. To this end, for each object in *S*, the object in *R* that has the highest number of corresponding pixels is identified and its label is saved in the list *assignments*, where this object might also be the background of *R*. Similarly, a list *reverse_assignments* is created, where now every object of *R* is assigned to an object of *S*. Then, the error counting metrics are defined as:

- *Split*: The number of objects in *R*, where two or more objects in *S* are assigned to only one object in *R*.

- *Merged*: The number of objects in *S*, where two or more objects in *R* are assigned to only one object in *S*.

- *Added*: The number of objects in *S* that are assigned to the background of *R*.

- *Missing*: The number of objects in *R* that are assigned to the background of *S*.

Given these definitions of the metrics, it is possible that a cell could be counted as missing, although it is only much smaller than it should be, because in this case the reference cell is assigned to the background. The other way round, cells with a considerably too large size might be identified as added objects.

5 DISCUSSION

If we want to compare algorithms for automatic segmentation in experimental data like fluorescent microscopy images of cells, we always face the problem that no actual ground truth is available. As a consequence, it is not a-priori clear what the 'right' result should look like. Thus, it is a common approach

Table 2: Summary of the quantitative evaluation measures computed for the segmentation results obtained by the various approaches considered in this paper (cf. also Figure 1 and Table 1).The left number always refers to the average value over all images of the respective measure for the U20S cell line, while the right number corresponds to the NIH3T3 cell line.

| pipeline	RI (in %)	JI	HD	NSD	split	merged	added	missing
manual	96.9 / 94.5	3.0 / 4.0	12.7 / 14.8	0.2 / 0.7	0 / 0	0 / 0.2	0.2 / 0.2	0 /2
variant 1	92.1 / 82.7	2.3 / 2.0	36.4 / 18.6	3.9 / 1.7	12.0 / 1.1	0.1 / 0.9	3.3 / 7.9	0.9 / 4.9
variant 2	91.6 / 82.7	2.2 / 2.0	36.7 / 18.1	4.0 / 1.7	12.4 / 1.1	0.1 / 0.9	4.1 / 8.1	0.9 / 4.8
variant 3	92.3 / 83.1	2.3 / 2.1	16.5 / 17.3	0.9 / 1.4	1.0 / 0.0	0.5 / 1.5	1.5 / 5.9	1.0 / 5.4
variant 4	97.0 / 86.5	2.7 / 2.3	13.9 / 16.2	0.6 / 1.0	1.0 / 0.0	0.5 / 1.6	1.5 / 5.4	0.5 / 4.0
variant 5	97.0 / 88.2	2.7 / 2.5	13.2 / 16.7	0.5 / 1.2	0.7 / 0.1	0.6 / 1.7	1.5 / 4.3	0.4 / 5.7
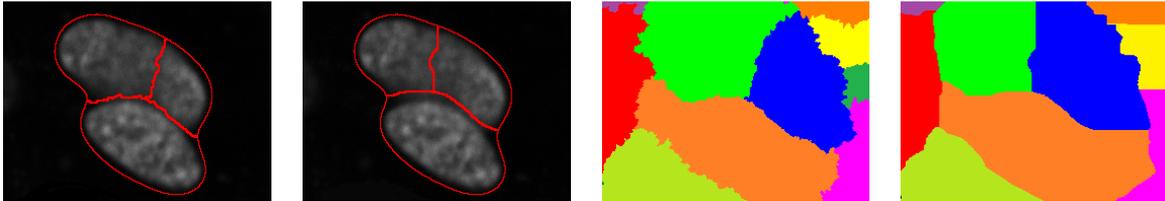
Figure 2: Close-up of the processed cell images. From left to right: Cell boundaries created with variant 1. Cell boundaries created with variant 2. Watershed transformation created with variant 1. Watershed transformation created with variant 2.

to rely on a manual 'expert' segmentation. However, since the latter is based on human experience, results may vary significantly between different manual segmentations (inter-observer variability), in particular since fluorescence microscopy images are typically characterised by low contrast, blur and other artefacts (cf. also (Sonka et al., 2014, p. 241)). These discrepancies are reflected by the numbers given in the first line of Table 2, which correspond to a comparison of the manually segmented images of A. Shariff and the respective manual segmentations of L. P. Coelho. In particular, these numbers indicate that we cannot expect any result obtained by an automatic segmentation routine to perfectly agree with the manual reference segmentation. After this preliminary remark, let us now briefly discuss the results of the various variants of watershed segmentation.

The first two variants of watershed segmentation we consider in this paper are based on the seeded watershed segmentation method described in (Coelho et al., 2009), where both the seed computation as well as the binary mask creation were carried out for a blurred version of the intensity image with $\sigma = 12$ as mentioned by the authors. The only difference between these versions is the image to which the watershed transformation (for fixed seeds) is applied. In the first case we apply it to the original intensity image, while in the second case a blurred version of the intensity image is used (cf. Table 1). We implemented both variants, since from our point of view the description in (Coelho et al., 2009) was not completely explicit in this respect, where regarding the evaluation in Table 2 the difference in the segmentation result

anyway seems to be marginal. This impression is further confirmed by the images in Figure 2. Here, the main difference appears to be that in the latter variant the watersheds separating the catchment basins follow a smoother course. Given these observations, we decided to not pursue the latter approach further but rather apply the watershed transform to the original intensity image to keep the complexity of the segmentation pipeline and the number of parameters as low as possible.

Looking again at the images in Figure 2, we moreover recognise that for both variants the upper cell in this close-up is split into two, an indicator that these variants of watershed segmentation have a tendency towards oversegmentation. Again this observation is in line with the numbers given in line two and three of Table 2, since the number of splits is far greater than the number of merged cells. Accordingly, we increased the value of σ (cf. number 1 in Figure 1), i.e. the amount of blur, which reduced the number of local minima and thus the number of seed points as illustrated by the two latter images in Figure 3. Note that we only increased the σ-value for the seed computation, but left the σ for the binary mask creation untouched. As can be seen in the fourth line of Table 2, this slight modification indeed resulted in a considerably reduced amount of split cells (an effect that is also reflected by the first two images in Figure 3), while on the other side of the coin the number of merged cells slightly increased. However, since the Rand Index and the Jaccard Index remained on a similar level, while the Hausdorff Distance as well as the NSD notably decreased, we conclude that all in all
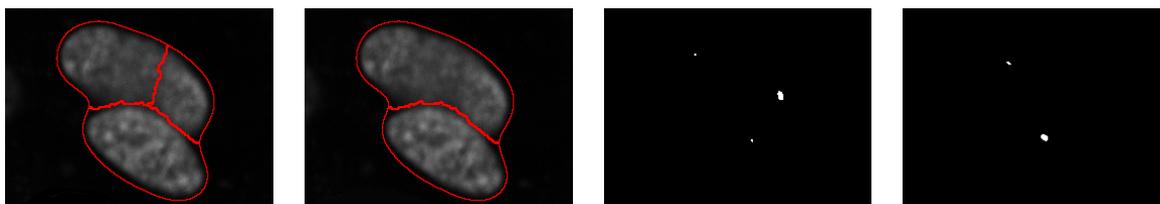
Figure 3: Close-up of the processed cell images. From left to right: Cell boundaries created with variant 1. Cell boundaries created with variant 3. Seed points created with variant 1. Seed points created with variant 3.
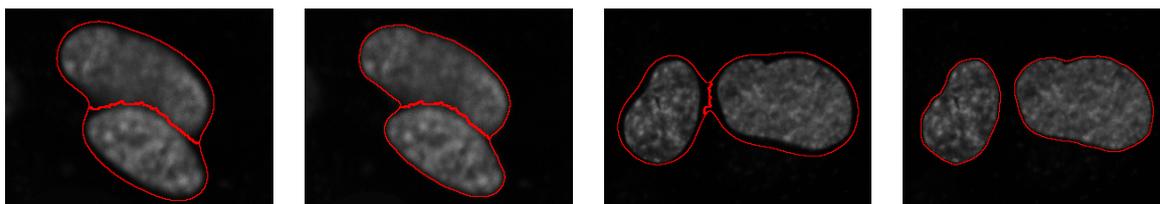

Figure 4: Close-up of the processed cell images. Left images: example of cell boundaries created with variant 3 and variant 4, respectively. Right images: another example of cell boundaries resulting from variant 3 and variant 4, respectively.
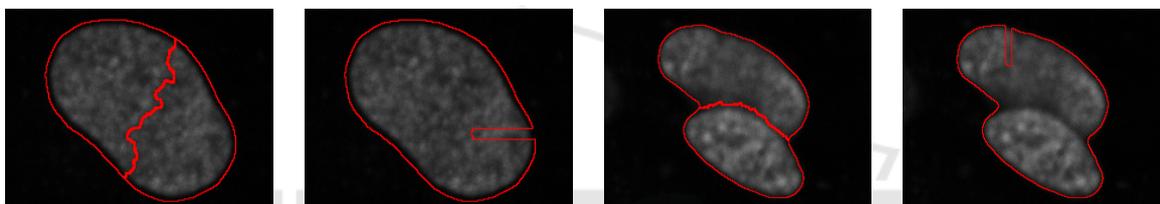

Figure 5: Close-up of the processed cell images. Left images: example of cell boundaries created with variant 4 and variant 5, respectively. Right images: another example of cell boundaries resulting from variant 4 and variant 5, respectively.

the reduction of seed points resulted in a better overall segmentation performance.

Taking another look at the second image in Figure 3, a remaining deficiency of the result provided by the third segmentation approach is that the red contour marking the detected object still encircles an area that is significantly larger than the actual cell. In order to address this issue, we conceived yet another version of watershed segmentation, where in comparison to the previous variant, we decreased the σ-value for the binary mask creation. Optimisation of this second σ-value just seemed to be the logical next step, since our segmentation pipeline anyway already included the computation of two different blurred versions of the original intensity image. As before, the optimisation of σ was carried out with respect to the first image and subsequently the segmentation pipeline with this parameter choice was applied to the entire sequence. In view of the images provided in Figure 4 and with regard to the numbers given in the fifth line of Table 2, we conclude that the latter version of watershed segmentation indeed yields rather convincing results for the given data set that could serve well as a baseline for the compara-

tive assessment of more recent and sophisticated fully automatic segmentation approaches.

The last version of watershed segmentation that we evaluated in this paper differs from the previously described version by the input of the watershed transform: the original intensity image is replaced by a distance transform of the latter. This version of seeded watershed has e.g. been employed in a more recent work of L. P. Coelho[7] and seems particularly suited for roundish objects, since in this case it might prevent oversegmentation as it is exemplified by the first two images in Figure 5. However, we also found cases like the one shown in the latter two images of Figure 5, where the distance transform-based watershed segmentation merged two cells that were separated by the previously discussed segmentation pipeline. With regard to the numbers given in Table 2, we see that all in all the evaluation measures indicate that on average the last two approaches seem to yield segmentation results of similar quality even though they both have their merits and demerits leaving room for

[7]https://github.com/luispedro/python-image-tutorial/blob/master/Segmenting%20cell%20images%20(fluorescent%20microscopy).ipynb

improvements by more advanced segmentation approaches. From our experience, we thus conclude that the latter two variants of watershed segmentation might both serve well as an evaluation baseline for the automatic segmentation of cell nuclei in fluorescence microscopy images by more involved methods.

6 CONCLUSIONS

In this paper we addressed the benchmarking of cell nuclei segmentation algorithms with a particular focus on fluorescence microscopy images. Specifically, we first described the considered dataset and explained how this data needs to be processed to serve our purposes, where we in particular pointed to several snares that might distort results if not properly taken care of. Afterwards, we recalled the watershed transformation and gave a detailed account about our implementation of a segmentation pipeline built around this well-known image decomposition tool. Next, we provided a review of three classes of well-established evaluation measures for image segmentation. Finally, we briefly compared the performance of several variants of our watershed segmentation pipeline, where we not only relied on the previously discussed quantitative evaluation measures but rather combined it with a visual inspection of the cell boundary images to collate the obtained results with our expectations. Everything combined, we thus explained the set-up of a watershed segmentation pipeline that might serve well as a baseline for the assessment of more sophisticated cell nuclei detection methods and as such constitutes an important component for our future efforts to contribute to a more personalised immune checkpoint inhibitor-based cancer therapy.

ACKNOWLEDGEMENTS

REFERENCES

Antonia, S. J., Villegas, A., et al. (2017). Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *New England Journal of Medicine*, 377(20):1919–1929.

Baltissen, D., Wollmann, T., et al. (2018). Comparison of segmentation methods for tissue microscopy images of glioblastoma cells. In *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*.

Bamford, P. (2003). Empirical comparison of cell segmentation algorithms using an annotated dataset. In *Proceedings of the 2003 International Conference on Image Processing (Cat. No.03CH37429)*. IEEE.

Chalana, V. and Kim, Y. (1997). A Methodology for Evaluation of Boundary Detection Algorithms on Medical Images. *IEEE Transactions on Medical Imaging*, 16(5):642–652.

Coelho, L. P., Shariff, A., and Murphy, R. F. (2009). Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In *Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE.

Digabel, H. and Lantuéjoul, C. (1978). Iterative algorithms. In *Proceedings of the 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, volume 19, page 8. Stuttgart, West Germany: Riederer Verlag.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579.

Kelly, P. N. (2018). The Cancer Immunotherapy Revolution. *Science*, 359(6382):1344–1345.

Osuna, E. G., Hua, J., et al. (2007). Large-Scale Automated Analysis of Location Patterns in Randomly Tagged 3T3 Cells. *Annals of Biomedical Engineering*, 35(6):1081–1087.

Peng, T., Bonamy, G. M. C., et al. (2010). Determining the distribution of probes between different subcellular locations through automated unmixing of subcellular patterns. *Proceedings of the National Academy of Sciences*, 107(7):2944–2949.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Russ, J. C. and Neal, F. B. (2015). *The Image Processing Handbook*. Taylor & Francis Inc.

Schadendorf, D., Hodi, F. S., et al. (2015). Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *Journal of Clinical Oncology*, 33(17):1889–1894.

Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image Processing, Analysis, and Machine Vision*. Cengage Learning.

Todaro, G. J. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *The Journal of Cell Biology*, 17(2):299–313.

Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598.