

## COMPARISON OF SEGMENTATION METHODS FOR TISSUE MICROSCOPY IMAGES OF GLIOBLASTOMA CELLS

D. Baltissen<sup>1</sup>, T. Wollmann<sup>1</sup>, M. Gunkel<sup>2</sup>, I. Chung<sup>3</sup>, H. Erfle<sup>2</sup>, K. Rippe<sup>3</sup>, K. Rohr<sup>1</sup>

<sup>1</sup> University of Heidelberg, BioQuant, IPMB, and DKFZ Heidelberg, Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group,  
<sup>2</sup> High-Content Analysis of the Cell (HiCell) and Advanced Biological Screening Facility, BioQuant, University of Heidelberg, Germany  
<sup>3</sup> Division of Chromatin Networks, DKFZ and BioQuant, Heidelberg, Germany

### ABSTRACT

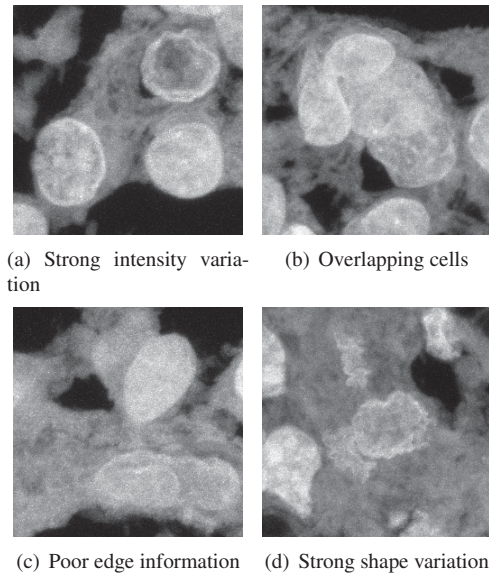
Segmentation of cell nuclei is crucial for analyzing microscopy image data and understanding cellular processes. We performed a comparison of different approaches for segmentation of cell nuclei in difficult tissue microscopy image data of glioblastoma cells. We investigated nine different methods comprising thresholding, deformable models, region growing, unsupervised learning, and supervised learning using random forest. In addition, we studied two deep learning methods including a novel deep neural network based on atrous spatial pyramid pooling (ASPP), which was not yet used for cell segmentation. We used pixel-based and object-based performance metrics. The best results were obtained by machine learning methods, where deep learning models (Accuracy 91.7% and Dice 0.925) exceeded random forest classifiers (Accuracy 90.4% and Dice 0.914). Using a deep neural network with an ASPP yielded a slightly lower segmentation performance, but was most effective in splitting merged objects.

**Index Terms**— Quantitative evaluation, segmentation, glioblastoma, fluorescence microscopy, machine learning

### 1. INTRODUCTION

The detection and segmentation of cells and cell nuclei constitutes an important task in microscopy image analysis. Challenges include image noise and other image artefacts such as image blurring. Further, cells or cell nuclei are heterogenous and vary in structure, shape, and intensity, or are clustering. In recent years, different methods for cell segmentation have been developed to address these challenges.

In this work, we investigated the performance of a wide range of different methods to segment nuclei from tissue microscopy images of glioblastoma cells. As can be seen in Fig. 1, this data is very challenging due to strong intensity variation, cell clustering, overlapping cells, poor edge information, missing object borders, strong shape variation,



**Fig. 1.** Examples of tissue microscopy images of glioblastoma cells with different challenges for image analysis.

and low signal-to-noise ratio. We compared nine different segmentation methods comprising thresholding, deformable models, region growing, unsupervised learning, and supervised learning. In previous work, different comparisons of methods for cell segmentation in fluorescence microscopy images were performed. Dima *et al.* [1] compared different segmentation methods using fluorescence microscopy data from two cell lines. The study revealed that K-means clustering yielded the best results, however, the used data does not seem to be very difficult and machine learning methods were not considered. Coelho *et al.* [2] evaluated segmentation algorithms using hand-labeled datasets including clustered nuclei. An approach based on merging multiple regions from watershed segmentation performed best, however, the focus of the study was on methods for high-throughput settings and therefore complex and time-intensive methods were not

included. Cheng and Rajapakse [3] presented a method to segment and separate clustered nuclei using shape markers in a watershed-like algorithm. The method was applied to noisy neuronal cell images. However, the used data is much less difficult compared to our data, particularly the image contrast at edges is much better. In Maška *et al.* [4] different methods for cell tracking were compared. The focus was on tracking, but the segmentation performance was also quantified. There, different types of data were considered compared to our study (e.g., human breast carcinoma cells, mouse embryonic stem cells, human squamous lung carcinoma cells).

In this contribution, we present a comprehensive comparison of nine different segmentation methods for difficult tissue microscopy data of glioblastoma cells (see Fig. 1). The comparison comprises a broad range of segmentation methods including thresholding, deformable models, region growing, unsupervised learning, and supervised learning using random forest. In addition, we studied two deep learning methods including a novel deep neural network based on atrous spatial pyramid pooling (ASPP), which was not yet used for cell segmentation. We provide quantitative results which reveal how well the different methods performed on the investigated challenging data.

## 2. METHODS

We studied nine different methods for segmentation of glioblastoma cells in tissue microscopy images.

### 2.1. Segmentation methods

*Global and local thresholding:* For global thresholding we used a maximum entropy approach (ImageJ Multithresholder plug-in), and for local thresholding the intensity contrast was employed (ImageJ auto local threshold plug-in).

*Fast marching and region competition:* For the fast marching algorithm [5], an image was first smoothed by a Gaussian filter and intensity maxima were used as seed points for a deformable model. For the region competition algorithm [6], we used a model based on a piecewise constant energy function and an energy length model based on a sphere regularization energy term. Local intensity maxima were used for initialization. The creation of handles and fission was enabled to cope with changes of the topology.

*Unsupervised learning by K-means clustering:* For K-means clustering the image was smoothed by a Gaussian filter and the number of clusters set to three (foreground, background, unspecific signal). The foreground cluster was used as segmentation result.

*Supervised learning using Weka and Ilastik:* Weka is a trainable segmentation tool [7]. We used a parallel random forest classifier containing 200 trees with unlimited depth. Feature selection was performed using the subset evaluation technique in conjunction with a greedy stepwise search

method (based on a J48-classifier), which resulted in the following features: Gaussian blur, Hessian, Membrane projections, Mean, Maximum, Anisotropic diffusion, Gabor, Laplacian, Entropy, Variance, Minimum, Median, Bilateral, Kuwahara, Structure and Neighbors. In addition, for supervised pixel classification we used a random forest classifier from Ilastik [8]. All provided features were used and the image scales were defined by  $\sigma=\{0.3, 0.7, 1.0, 1.6, 3.5, 5.0, 10.0\}$ . Both classifiers (Weka, Ilastik) were trained using 20 images.

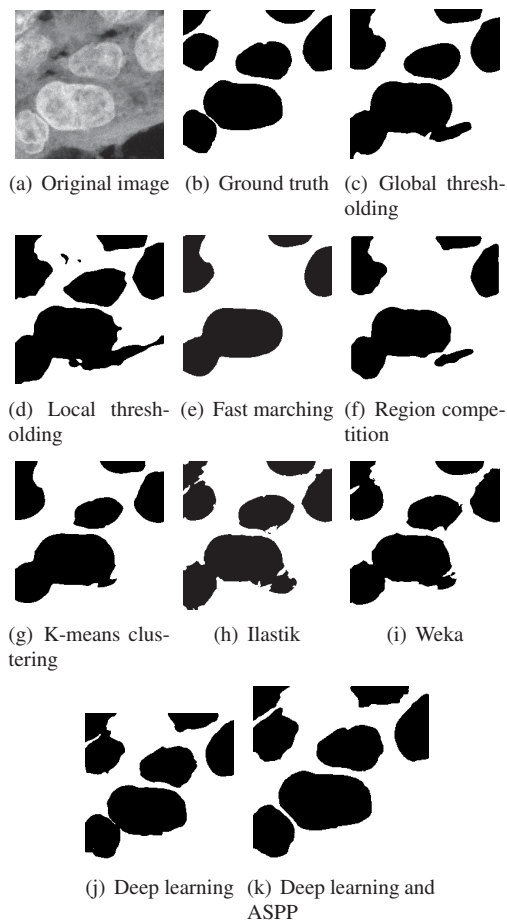
*Deep learning:* We used two deep neural networks based on U-Net [9]. The first network is similar to a vanilla U-Net using batch normalisation and residual connections. The second network combines a U-Net with atrous spatial pyramid pooling (ASPP) [10]. ASPP has the advantage that image context can be captured at multiple scales. Using a U-Net in conjunction with a PReLU activation function, we observed that the first layers mostly favour negative activations. However, PReLU reduces the execution time. Therefore, we used PReLU only in the first layer to make use of negatively activated features, while saving execution time. For the second network, we used an ASPP block after the last residual block in the downsampling path followed by Gaussian dropout ( $p = 0.5$ ). Compared to the original ASPP we used dilations of 1, 2 and 4, and global average pooling (pooling kernel equal to feature maps). The extracted feature maps are concatenated and fed to a  $1 \times 1$  convolution for compression. Data augmentation was performed using random flipping, rotation, scaling, noise addition, and edge aware elastic deformation. The network was trained on  $128 \times 128$  pixel patches with a balanced foreground and background ratio. We trained the network using cross-validation and early stopping with Adam as optimizer.

### 2.2. Performance measures

The performance of the different methods was determined by comparing the segmentation results with ground truth data generated by manually outlining nuclei. We employed multiple well established pixel-based performance measures, namely the Dice coefficient (Dice), Jaccard index (JI), Rand index (RI), Hausdorff distance (HD), Sensitivity (Sens.), Specificity (Spec.), and Accuracy (Acc.). The performance values were determined using the Visceral EvaluateSegmentation tool. For a detailed description of the performance measures we refer to [11]. In addition, we determined object-based performance measures. We used the Jaccard index (JI) at object level and also determined the number of missing, erroneously added, and incorrectly merged objects.

## 3. EXPERIMENTAL RESULTS

We have applied all nine segmentation methods to 50 fluorescence microscopy tissue images of glioblastoma cells. The



**Fig. 2.** Segmentation results of different methods for an example of a tissue microscopy image of glioblastoma cells.

images were acquired by a Leica TCS SP5 point scanning confocal microscope with a 63x objective lens. The voxel size was 100 nm in the  $xy$ -plane and 250 nm in  $z$ -direction. 45 axial layers were acquired for each stack of the DAPI channel by exciting with a violet (405 nm) laser. For segmentation, maximum intensity projection (MIP) images the stacks were used. Example images are shown in Fig. 1. The dataset was split into 30 images for hyperparameter optimization and 20 images for performance evaluation. Example segmentation results of the investigated methods are shown in Fig. 2. It can be seen that both fast marching and region competition missed to segment one cell nucleus in the middle and close objects were mostly merged. Only deep learning with ASPP achieved a clear separation of close-by objects.

Quantitative results of all nine segmentation methods using pixel-based and object-based performance measures are provided in Table 1. For the pixel-based metrics it can be seen that the two thresholding methods generally yielded the lowest values. For the machine learning methods we obtained better results for Specificity and Sensitivity. Fast marching yielded the best result for HD. The deep learning model with

out ASPP achieved the best values for Dice (0.925), Jaccard index (0.866), Rand index (0.853), Sensitivity (0.953), Specificity (0.871), and Accuracy (0.917). The best random forest classifier (Weka) was somewhat worse (Dice: 0.914, Accuracy: 0.904). Deep learning with ASPP yielded slightly lower values (Dice: 0.911, Accuracy: 0.901).

The results for the object-based metrics show that both deep learning models yielded an above-average result for the number of erroneously added objects of 15 (without ASPP) and 23 (with ASPP). The deep learning model with ASPP showed a higher object-based Jaccard index (0.833) compared to the model without ASPP (0.770). Concerning object merging, region competition yielded 36 merged objects, which is better compared to K-means clustering (44) and fast marching (42), although three objects were not detected. Both deep learning models merged only few objects incorrectly. The best result among all methods was achieved for the model with ASPP (14), the second best result was achieved for the model without ASPP (23). Thus, deep learning with ASPP was most effective in splitting merged cells.

#### 4. DISCUSSION AND CONCLUSION

We presented a comprehensive performance evaluation of nine different segmentation methods applied to challenging microscopy tissue images of glioblastoma cells. Each method showed specific strengths and weaknesses, which are generally related to the underlying model and assumptions.

Although the two thresholding methods yielded the worst results, their application is very user-friendly and processing is very fast. Both fast marching and region competition yielded slightly better results, however, the performance was limited due to the poor edge information in the image data which resulted in a tendency to over-segment some objects. On the other hand, both methods require seed point initialisation. In this study, detected intensity maxima were used. As a result, for relatively dark objects partly no seed points were detected and the objects were not segmented. Furthermore, due to the large variation of the image data, we had to use two different parameter settings for fast marching to obtain acceptable results.

The K-means algorithm requires setting only a small number of parameter, and application is easy and fast. The obtained results were better than thresholding, fast marching, and region competition. A problem of K-means clustering is that the cluster centers are randomly initialized and therefore different results are obtained depending on the initialization. Furthermore, the result depends on the chosen number of clusters. We obtained the best results for three clusters.

Further, we studied supervised learning methods based on random forest classifiers (Weka and Ilastik). Weka provides a wider range of image features and classifiers compared to Ilastik. For both tools, orientation features (Hessian) and texture or structure features were more effective compared to

Method	Pixel-based metrics							Object-based metrics			
	Dice	Jl	RI	HD	Sens.	Spec.	Acc.	Jl	Missing	Added	Merged
Global thresholding	0.884	0.807	0.794	1.665	0.928	0.792	0.864	0.582	1	25	43
Local thresholding	0.881	0.792	0.773	1.139	0.893	0.823	0.864	0.480	2	14	50
Fast marching*	0.905	0.832	0.814	<b>0.775</b>	0.933	0.836	0.891	0.491	3	<b>5</b>	42
Region competition	0.904	0.829	0.810	0.986	0.934	0.828	0.890	0.546	3	9	36
K-means clustering	0.910	0.839	0.821	0.848	0.927	0.846	0.896	0.531	1	8	44
Ilastik	0.911	0.845	0.828	0.794	0.941	0.841	0.897	0.610	<b>0</b>	11	34
Weka	0.914	0.848	0.833	0.814	0.939	0.853	0.904	0.570	1	8	36
Deep learning	<b>0.925</b>	<b>0.866</b>	<b>0.853</b>	1.102	<b>0.953</b>	<b>0.871</b>	<b>0.917</b>	0.770	<b>0</b>	15	23
Deep learning and ASPP	0.911	0.843	0.829	1.299	0.946	0.847	0.901	<b>0.833</b>	<b>0</b>	23	<b>14</b>

**Table 1.** Pixel-based (left) and object-based (right) performance metrics for different segmentation methods. The values are mean values over 20 images. The best results are highlighted in bold. \* Two different parameter settings used

common intensity and edge features. Hence, both classifiers yielded better results than thresholding, fast marching, region competition, and K-means clustering.

The two deep learning models showed the best results, although training is computationally expensive. For the model with ASPP we used an ASPP block after the last residual block in the downsampling path which allows exploitation of large context information. Although the pixel-based performance of the deep learning model with ASPP was generally slightly lower compared to the model without ASPP, the model with ASPP showed its strength in separating merged objects. Compared to all other methods, this model yielded the best result with 14 merged objects which is only 39 % of the average number of merged objects for all methods (36), compared to 64 % (deep learning without ASPP) and 94 % (Ilastik). A further improvement of the deep learning models could be obtained by increasing the number of images for training.

Nevertheless, all methods yielded merged objects and therefore an additional post-processing step is generally required to split these objects (e.g., using a watershed transform). A disadvantage of both deep learning models is that they are based on supervised learning and therefore ground truth data is required. Ground truth has to be created manually, which is a tedious task and should be performed by experts.

In summary, the deep learning models performed best although training is very time consuming. The deep learning model without ASPP outperformed all other tested methods for nearly all pixel-based performance metrics. The only method that yielded better results for the object-based metrics was deep learning with ASPP. Deep learning with ASPP showed best results for the separation of clustered cell nuclei.

### Acknowledgements

Support of the BMBF within the projects CancerTelSys (e:Med), RNA-Code (e:Bio), and de.NBI (HD-HuB) is gratefully acknowledged.

### 5. REFERENCES

- [1] A. A. Dima, J. T. Elliott, J. J. Filliben, M. Halter, A. Peskin, J. Bernal, M. Kociolek, M. C. Brady, H. C. Tang, and A. L. Plant, "Comparison of segmentation algorithms for fluorescence microscopy images of cells," *Cytometry Part A*, vol. 79, no. 7, pp. 545–559, 2011.
- [2] L. P. Coelho, A. Shariff, and R. F. Murphy, "Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms," in *Proc. ISBI 2009*, pp. 518–521, IEEE, 2009.
- [3] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 741–748, 2009.
- [4] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak, *et al.*, "A benchmark for comparison of cell tracking algorithms," *Bioinformatics*, vol. 30, no. 11, pp. 1609–1617, 2014.
- [5] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, no. 4, pp. 1591–1595, 1996.
- [6] J. Cardinale, G. Paul, and I. F. Sbalzarini, "Discrete region competition for unknown numbers of connected regions," *IEEE Trans. Image Proc.*, vol. 21, no. 8, pp. 3531–3545, 2012.
- [7] I. Arganda-Carreras, V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, and H. S. Seung, "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification," *Bioinformatics*, pp. 2424–2426, 2017.
- [8] C. Sommer, C. Straehle, U. Köthe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *Proc. ISBI 2011*, pp. 230–233, IEEE, 2011.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI 2015*, pp. 234–241, Springer, 2015.
- [10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *arXiv:1606.00915*, 2016.
- [11] E. D. Gelasca, J. Byun, B. Obara, and B. Manjunath, "Evaluation and benchmark for biological image segmentation," in *Proc. ICIP 2008*, pp. 1816–1819, IEEE, 2008.