

GLOBAL PROBABILISTIC MODELS FOR ENHANCING SEGMENTATION WITH CONVOLUTIONAL NETWORKS

Mengran Fan and Jens Rittscher

Institute of Biomedical Engineering, University of Oxford, Oxford, UK

ABSTRACT

While deep learning has dramatically improved our capabilities for developing extremely robust segmentation methods, some challenges remain. In many practical settings we only have access to a limited amount of training data. More importantly, the relationship between algorithm performance and the required amount of training data is not well understood.

Here we propose to combine convolutional network based segmentation approaches with a global probabilistic model that effectively enforces prior shape constraints. We demonstrate that the model is capable of accurately segmenting densely packed populations of cells. Our experiments show that combining the convolutional network with the proposed model-based segmentation approach improves the overall segmentation accuracy.

Index Terms— High-content imaging, convolutional networks, segmentation, fluorescence microscopy.

1. INTRODUCTION

The robust and reliable delineation of individual objects in a densely packed population of cells is a fundamental challenge of live cell imaging. While certain sample preprocessing methods can be utilised to control the overall plating density of cells, it is very difficult to avoid their aggregation in realistic experimental settings. As the field progresses towards developing more complex in vitro cell models that promise to be more phenotypically relevant and predictive, there is an increasing need for principled solutions.

Recently published deep learning approaches, in particular those that have been tailored to biomedical applications [1, 2], are very successful. However, our experiments demonstrate that when trained on smaller data sets, such models have problems resolving densely packed cells (see Figure 1). We propose to complement the convolutional neural network (CNN) based architectures with a more traditional model-based approach. In particular, we make use of an algorithm [3] that has been developed for segmenting groups of people in surveillance videos. Here, certain assumptions about the shape of each individual object are utilised to segment objects by grouping a set of image features in a globally consistent manner.

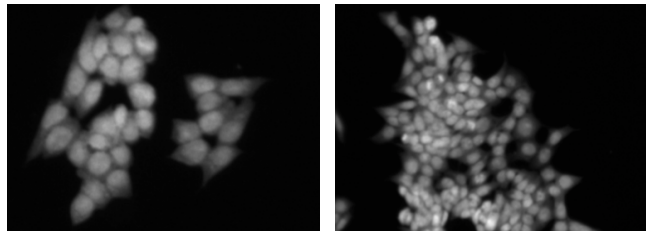


Fig. 1. Challenges in realistic experimental settings. *The dense packing of the cells makes their segmentation difficult. In addition, we observe that the fluorescent dye bleeds into the cytoplasm of the cells.*

We demonstrate that a combination of these techniques leads to significant improvements in segmentation accuracy. While our current work focuses on segmenting monolayers of cells, it is clear that this approach can be extended to 3D and can also be applied to more complex cell types. First we establish a baseline by applying a U-Net [1] type CNN to our data set. The associated failure modes are being discussed in Section 3. The mathematical framework of our model-based segmentation framework are being presented in Section 4. Selected examples that illustrate how the proposed model can delineate densely packed cells are presented in Section 6. In addition, we present selected examples that illustrate how the proposed model can delineate densely packed cells. Finally, we present a summary of our contributions and outline future work.

2. RELATED LITERATURE

Lin *et al.* [4] extend the standard watershed segmentation with a gradient-weighted distance function. This approach effectively identifies inflection points to split pairs of touching cells. In a similar fashion, He *et al.* [5] combine the localisation of intervening and concave contours with a normalised cut algorithm to achieve accurate segmentations.

Using standard neural networks Nandy and collaborators [6] develop an approach that uses supervised learning to effectively break irregularly large segments into valid objects. Their approach uses multiple watershed segmentations and a tree based merging approach to achieve these results.

Mosaliganti *et al.* [7] address the problem of segmenting densely packed cells in tissues using level sets. They propose an effective way to learn the model parameters from training data. Bergeest and Rohr [8] proposed to combine three different energy functionals. While the approach effectively combines local and global information the resulting framework is complex.

Van Valen *et al.* [2] report a number of successful experiments that involve the segmentation of cells from various different cell lines with CNNs. However, they note that use the output obtained from these CNNs to guide active contours to obtain accurate segmentations. In a more general segmentation setting similar observations can be made. Chen *et al.* [9] integrate a conditional random field in the network to improve the segmentation accuracy. They use active contours to process the output generated by CNNs. A similar approach can be used to improve the pixel level accuracy of brain tumour segmentations [10].

As U-Net is widely recognised as one of the most popular CNN architectures for biomedical segmentation, we use it to establish baseline performance. In the following section we discuss the application of U-Net to our particular data set and motivate how a model-based framework can be used to improve the performance when only limited training data is available.

3. DEEP LEARNING BASED SEGMENTATION

To train U-Net, we annotated 1900 cells in 30 images of our data set, which consists of a high-throughput experiment using haploid cells. Cell nuclei are stained with DAPI. The images in Figure 1 indicate that the nuclear marker can bleed into the cytoplasm of these cells which causes additional visual clutter. The overall goal of this experiment is to study the morphological variation of these cells under various different conditions which motivates the need for an accurate delineation of individual cells.

In the training phase, we randomly extracted 128 x 128 patches from 30 annotated images of size 1024 x 1360. This way the amount of training data is being increased. The network architecture is similar to the original structure proposed in [1]. Considering the size of cellular objects, we only employ the typical architecture of convolution(+ReLU)-convolution(+ReLU)-max pooling twice in the downsampling path. The corresponding upsampling path also consists of two repeated application of upsampling-convolution(+ReLU)-convolution(+ReLU). Our model is relatively shallow as segmenting individual cells does not require large receptive fields. Furthermore, we used back-propagation and stochastic gradient descent for optimization.

In practice, two separate networks were trained in our experiments, namely U-Net-A and U-Net-B. All classes had equal weights in the overall loss function. Based on the segmentation results, it can be concluded that U-Net-A can ef-

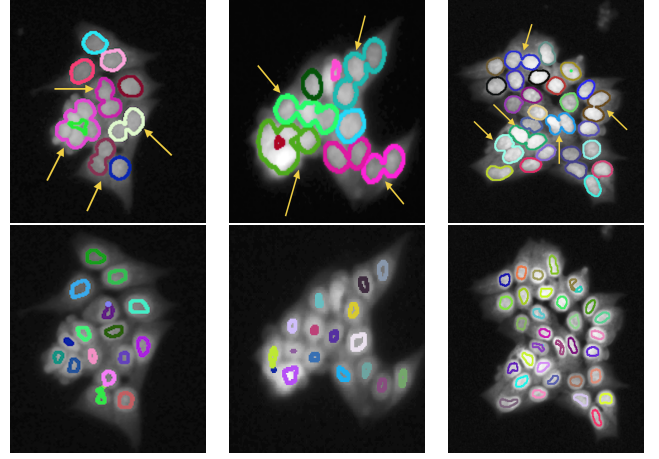


Fig. 2. Failure modes of U-Net based segmentation. The first row presents the result of U-Net-A which is trained using an equally weighted loss function. It can be observed that the network identifies a clump of cells as one object. The second row shows the failure modes of U-Net-B which use a weighted loss function. Although it can separate the clustered nuclei successfully, the delineation of the object boundaries is not very accurate.

fectively capture the isolated nuclei, whereas cell clumps are typically recognised single objects. Alternatively, we used a weighted loss function in this network to enhance the boundaries of cells. Comparing with U-Net-A, U-Net-B tends to separate the clustered nuclei successfully. However, most pixels in the interior and the cytoplasm neighbouring the boundaries were also predicted as boundaries in the testing phase, leading to a low pixel-level accuracy. The examples of failure modes of U-Net-A and U-Net-B are illustrated in Figure 2.

Our experiments indicate that deep learning based methods are not sufficient to localise the accurate boundaries for individual cells in pixel level, especially when the cells are tightly packed. This is mainly due to the build-in invariance that makes deep learning networks desirable for high-level vision tasks rather than delineating outlines of cellular structures.

Although one could make use of additional training data or data augmentation to overcome this problem, it is important to design robust algorithms on smaller training data sets. We will demonstrate later that our global probabilistic model, which is introduced in the next section, can be used to overcome these failure modes.

4. MODEL-BASED SEGMENTATION

In this section we describe our model-based segmentation approach which will be used as a final step to obtain a globally optimal segmentation. Core principle of model-based approach is to assign image observations \mathcal{Z} to a set of object

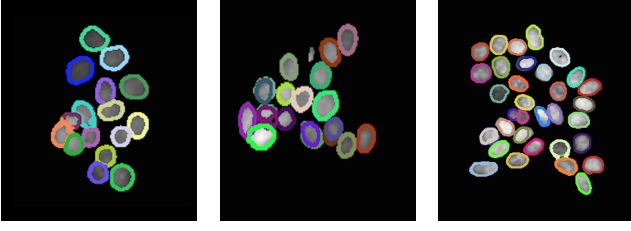


Fig. 3. Model-based segmentation. The figure shows selected results of our model-based segmentation methods that correspond to failure modes that were highlighted in Figure 2

hypotheses \mathcal{C} that are consistent with a parametric prior shape model. In our case we utilise SLIC [11] to generate a set of superpixels, which encodes various low level image features. Learnt edge profiles are used as an additional source of information. The association of image features $Z = \{z_i\}$ and object hypotheses \mathcal{C} is captured with the help of an assignment vector $V = \{z_i\}$. We explicitly model the joint probability of an assignment vector V and a feature set Z as $P(V, Z|X)$, where X acts as a parameter of the distribution. Crucially, X captures the shape of each object in the scene. As the assignments V cannot be observed directly, we treat V as a hidden parameter. Prior work motivates that we can estimate X by optimising the log likelihood $L(V|Z; X)$ of a feature assignment V given a set of image features Z . Here it is important to note that this approach is an effective global optimisation that takes all the information in the image into account.

The joint probability model $P(V, Z; X)$ is based on a set of merit functions that measure the affinity of a single or a pair of image features with a given hypothesis. Here we capture the affinities of single feature assignments and as well as pairwise assignments to a given potential hypothesis h_k with parameter x_k . The feature assignment V is represented as a matrix $M = M_{ik}$ where m_{ik} is the probability of i^{th} image feature z_i belonging to the k^{th} hypothesis h_k . We formulate the log likelihood of a feature assignment V given a set of image feature Z as:

$$L(V|Z; X) \propto \gamma_1 \sum_{k=1}^K \sum_{i=1}^K g(z_i, x_k) m_{ik} + \gamma_2 \sum_{k=1}^K \sum_{i,j=1; i \neq j}^K g(z_i, z_j, x_k) m_{ik} m_{jk}, \quad (1)$$

where the $g(z_i, x_k)$ and $g(z_i, z_j, x_k)$ denote the merit or affinity functions. As the value of $p(Z)$ remains constant throughout the whole process, the joint probability can be written as

$$p(V, Z; X) = p(V|Z; X)p(Z) \propto \exp(L(V|Z; X)). \quad (2)$$

Rittscher *et al.* [3] demonstrate that the Maximum-a-Posteriori (MAP) estimate of X can be found by a variant of the Expectation Maximisation which iteratively optimizes the

free energy

$$F(\tilde{P}, X) = E_{\tilde{P}}[\log P(V, Z; X)] + H(\tilde{P}), \quad (3)$$

where \tilde{P} is the probability density function of V , and H denotes the entropy of this distribution. The EM algorithm can be expressed as follows:

$$\begin{aligned} \text{E-step: Set } \tilde{P} \text{ to the } \tilde{P} \text{ that maximises } F(\tilde{P}, X) \\ \text{M-step: Set } X^* \text{ to the } X \text{ that maximises } F(\tilde{P}, X). \end{aligned} \quad (4)$$

The available space does not allow us to present the implementation of this optimization procedure in detail. In practice it is implemented through an annealing process. An explicit regularity parameter T is added to the optimization function to globally control the probability distribution:

$$F(\tilde{P}, X) = E_{\tilde{P}}[\log P(V, Z; X)] + TH(\tilde{P}), \quad (5)$$

In addition the affinity functions g is used to model this joint likelihood probability. To make it robust against various different image conditions, both shape and textual information are encoded in this model. Any given object hypothesis x_k specifies an ellipse with an intensity pattern. Given a single superpixel, its boundary and intensity distribution are used to evaluate how well it fits to this specific object hypothesis. Here, if two superpixels have very similar internal texture, it is likely that they belong to the same object. In case two superpixels are divided by a strong edge it is unlikely that they belong to the same object. The parameters of the affinity functions g are optimized on a set of example image patches. Finally, with a modified version of EM algorithm, nuclei hypotheses are distinguished from other invalid hypotheses.

5. OVERALL SEGMENTATION FRAMEWORK

We now discuss how the proposed model-based approach can be utilised to further improve the results obtained by convolutional networks. In our setting, very basic constraints on shape and size are being used to detect outliers. Segments which result in a low ellipse fitting score or are simply too large are being processed further.

As indicated in Figure 4, subsequently superpixels are generated for the corresponding image regions. After feature extraction the stochastic optimisation described in Section 4 is being applied to the local image patch. The object boundaries proposed by U-Net are used as proposed edge maps in the model-based segmentation. Here, inflection points are detected as potential locations for splitting touching cells. The overall segmentation algorithm is shown in Figure 5

We argue that the proposed framework is a very natural and efficient way of combining a data driven and a model-based approach. The convolutional network robustly identifies isolated cells and adapts very efficiently to the given imaging conditions. In contrast, the model-based approach effectively

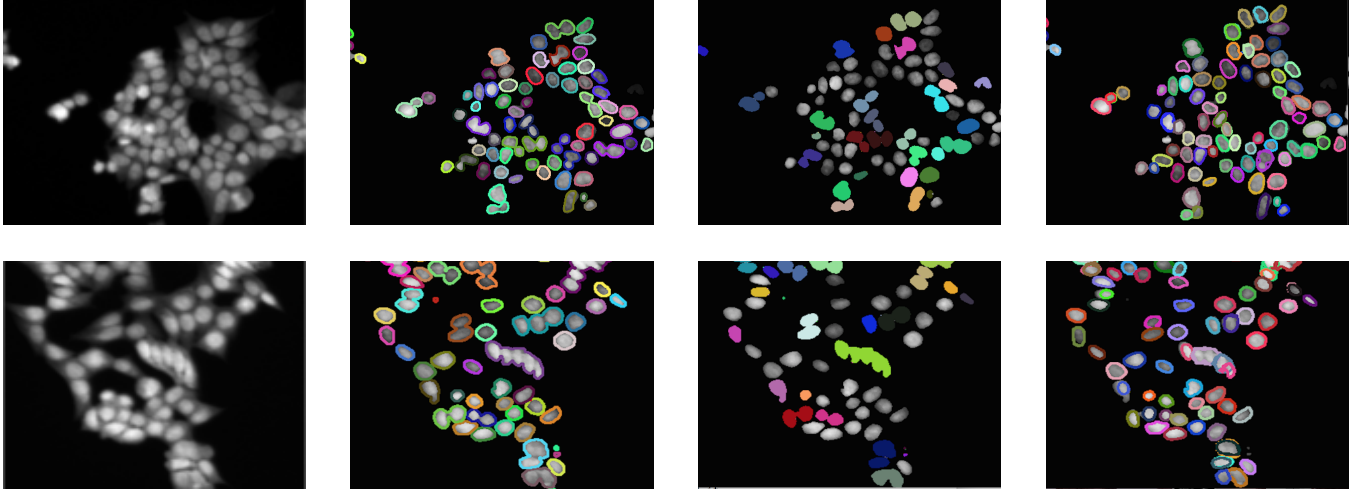


Fig. 4. Step-wise result for two sample images. The input images are shown in the first column. The second column presents the segmentation results of U-Net model. Based on the simple shape and size constraints, the third column describes the result of outlier detection. Finally, the segmentation results after globally probabilistic optimisation are shown in the last column.

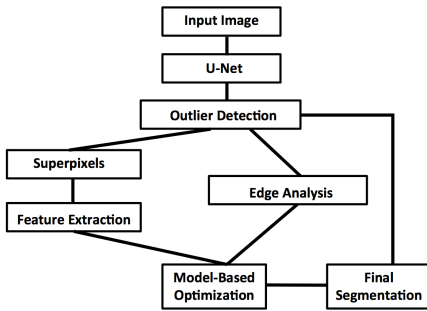


Fig. 5. Overall segmentation algorithm. See Section 5 for details.

enforces some basic shape constraints. Local appearance and spatial consistency are explicitly captured in this model with the help of the affinity functions g .

6. EXPERIMENTAL RESULTS

The main results of our experiments are summarised in Table 1. We performed segmentation of nuclei on 20 images containing about 1500 nuclei in total. As we employ a relatively small training set, U-Net-A tends to give the wrong prediction for clustered nuclei. The weighted loss function can be adopted to address the problem of cell clumps in high level vision tasks like cell counting. However, considering the pixel level segmentation accuracy, the pixels in cell interior and cytoplasm tend to be predicted as edge points by U-Net-B. Finally, when the annotated training data is limited,

our global model-based approach effectively compensates for potential segmentation failures. Table 1 shows this framework achieved 9% improvement over the existing benchmark segmentation approaches.

Method	Jaccard Index	Dice Index
U-Net-A	0.67	0.80
U-Net-B	0.72	0.84
U-Net+Probabilistic model	0.81	0.90

Table 1. Comparison of segmentation performance.

7. SUMMARY & CONCLUSIONS

Deep learning opens up new and exiting opportunities for semantic segmentation. However, in many practical applications only a limited amount of training data is available. Here, we demonstrate that a global probabilistic model can be used to effectively compensate for the shortcomings of segmentations that have been obtained by a CNN. In the future we plan a tighter integration of these two approaches. For example, the image features learnt from the convolutional networks should be used for generating the superpixels. Our experiments already indicate that this would further improve our results.

8. ACKNOWLEDGEMENTS

JR is supported by the EPSRC SeeBiByte Programme Grant (EP/M013774/1) and the Ludwig Institute for Cancer Research.

9. REFERENCES

- [1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” may 2015.
- [2] David A Van Valen, Takamasa Kudo, Keara M Lane, Derek N Macklin, Nicolas T Quach, Mialy M DeFelice, Inbal Maayan, Yu Tanouchi, Euan A Ashley, and Markus W Covert, “Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments,” *PLoS computational biology*, vol. 12, no. 11, pp. e1005177, 2016.
- [3] Jens Rittscher, Peter H Tu, and Nils Krahnstoeber, “Simultaneous estimation of segmentation and shape,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 486–493.
- [4] G Lin, U Adiga, K Olson, J F Guzowski, C A Barnes, and B Roysam, “A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks,” *Cytometry Part A*, vol. 56, no. 1, pp. 23–36, 2003.
- [5] Yong He, Hui Gong, Benyi Xiong, Xiaofeng Xu, Anan Li, Tao Jiang, Qingtao Sun, Simin Wang, Qingming Luo, and Shangbin Chen, “icut: an integrative cut algorithm enables accurate segmentation of touching cells,” *Scientific reports*, vol. 5, pp. 1–17, 2015.
- [6] Kaustav Nandy, Prabhakar R Gudla, Ryan Amundsen, Karen J Meaburn, Tom Misteli, and Stephen J Lockett, “Supervised learning framework for screening nuclei in tissue sections,” *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2011, pp. 5989–92, jan 2011.
- [7] Kishore Mosaliganti, Arnaud Gelas, Alexandre Gouaillard, Ramil Noche, Nikolaus Obholzer, and Sean Megason, “Detection of spatially correlated objects in 3d images using appearance models and coupled active contours,” *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 641–648, 2009.
- [8] Jan-Philip Bergeest and Karl Rohr, “Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals,” *Medical Image Analysis*, vol. 16, no. 7, pp. 1436–1444, 2012.
- [9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *arXiv preprint arXiv:1606.00915*, 2016.
- [10] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan, “A deep learning model integrating fc-nns and crfs for brain tumor segmentation,” *arXiv preprint arXiv:1702.04528*, 2017.
- [11] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.