CNN Based Yeast Cell Segmentation in Multi-Modal Fluorescent Microscopy Data

Ali Selman Aydin *[1], Abhinandan Dubey[1], Daniel Dovrat[2], Amir Aharoni[2] and Roy Shilkrot[1]

[1]Department of Computer Science, Stony Brook University, USA
[2]Department of Life Sciences and the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel

Abstract

We present a method for foreground segmentation of yeast cells in the presence of high-noise induced by intentional low illumination, where traditional approaches (e.g., threshold-based methods, specialized cell-segmentation methods) fail. To deal with these harsh conditions, we use a fully-convolutional semantic segmentation network based on the SegNet[3] architecture. Our model is capable of segmenting patches extracted from yeast live-cell experiments with a mIOU score of 0.71 *on unseen patches drawn from independent experiments. Further, we show that simultaneous multi-modal observations of bio-fluorescent markers can result in better segmentation performance than the DIC*[1] *channel alone.*

1. Introduction

Accurate segmentation in cellular microscopy is a vital step for successful analysis of experiments in living cells. Any level of automation of this process is a boon for researchers, as manual analysis of cell shapes is highly laborious and requires expert knowledge of the biological constructs as well as the imaging equipment. Hence, various techniques were developed and employed for this purpose. Traditional techniques for cell segmentation in 2D digital microscopy were proposed as early as the 1960s[11], starting with more fundamental methods such as thresholding, the watershed transform [13] and deformable models [4]. Software packages tailored for the task of cell segmentation soon followed [6, 12, 19].

Many of the existing cell segmentation algorithms assume a relatively clean background with near-identical and easily differentiable foreground cell shapes. Such assumptions make it possible to effectively utilize methods like thresholding, region growing or watershed transform with tweaking of global parameters. However, in some cases the microscopy imaging of a live-cell experiment can be noisy for reasons related to the constraints of the experiment. One example, is the case where illumination of the sample should be minimized in order to maintain cell viability and prevent DNA photo-damage caused by excessive illumination. Low illumination is particularly important for work involving DNA replication, since high-intensity lighting can easily damage DNA, and disrupt or bias the experiment results. One example of such an experiment is shown in Figure 4. Notice that the cells of interest come in all shapes, brightness, textures and transparency levels, in contrast to many of the live-cell experiments available to the community. In certain cases, the cells are virtually indistinguishable from the background, and human annotators cannot achieve consensus whether a certain region belongs to a cell or not. To illustrate this, we computed the match between two human annotators to be only 79% in our dataset (In terms of mIOU, please refer to Equation 1 for details).

The nature of the experiments, like the one mentioned above, makes it hard for many existing algorithms to correctly segment the cell regions. Therefore more powerful methods are needed in order to distinguish the regions of the cell from the background under challenging conditions.

In this work, we try to address the aforementioned difficulties of the segmentation task in low illumination scenarios. We demonstrate that a CNN-based segmentation method (see Section 3) outperforms many widely-used methods applied to our unique dataset (see Section 4), and is able to cope with high noise and significant variation in cell shapes and appearances (see Section 5).

Contributions in this paper are as follows:

- We demonstrate that using deep learning semantic seg-

[*] aaydin@cs.stonybrook.edu
[1] Differential Interference Contrast

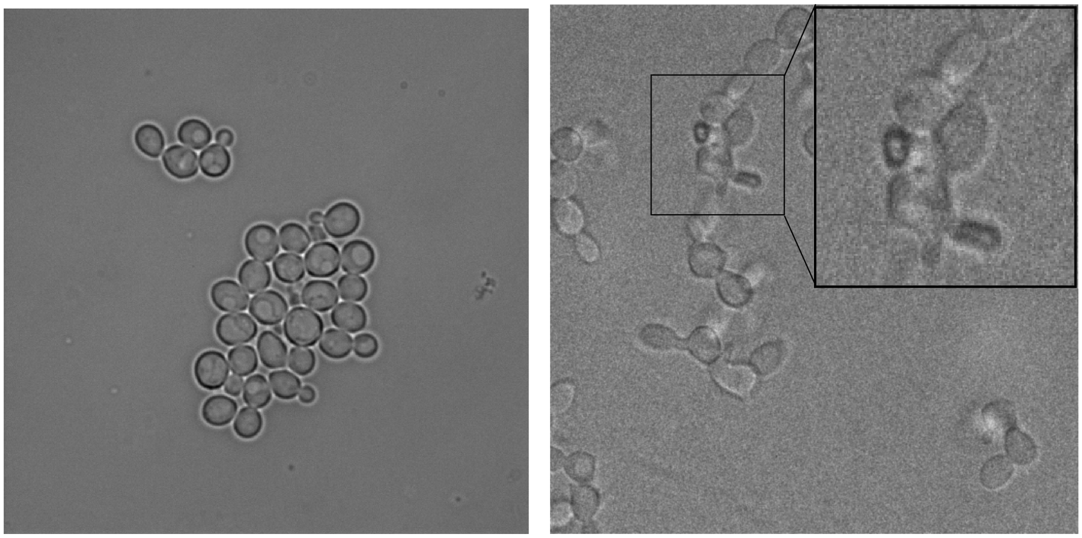


Figure 1: Comparison of a usual yeast image dataset (left) [17] and our dataset (right). Notice the high noise and variation of color and texture present in our dataset, compared to the dataset on the left.

mentation results in a significant improvement over existing approaches when the experimental settings introduce non-trivial conditions (*e.g.* noise, high variation in color/intensity).

- We hypothesize that multi-modal imaging of additional fluorescent microscopy channels can be useful in improving segmentation performance compared to only the live-cell video. We provide empirical evidence for this by showing that a small improvement in mIOU can be achieved simply by feeding extra channels to the segmentation network.

2. Related Work

It is possible to group the existing approaches for cell image segmentation into three overarching themes of computer vision paradigms.

Unsupervised methods: such as thresholding [2], active contours [5], watershed, etc.

Supervised methods: for example [10] that uses detected image boundaries to train a set of classifiers, which may be employed to detect cell boundaries. A hybrid approach has been proposed in [9], which starts with a supervised algorithm to convert color images to grayscale, followed by histogram thresholding and watershed based segmentation.

Manual parametric methods that allow the user to combine the existing approaches and experiment with the parameters in a hands-on manner. These solutions are usually provided as easy-to-use standalone software or plugins. Some examples for such software packages are CellProfiler [6], Cellstar [19] and Outfi [12].

Similar to other vision problems, Deep Learning approaches recently became popular for cell segmentation problems and biomedical imaging segmentation problems at large [18, 14]. DeepCell [18] is particularly noteworthy, since it shares the focus of our work of live-cell imaging experiments. The authors of DeepCell have re-formulated the cell image segmentation problem as an image classification problem, where small patches on a sliding window are extracted from the image and classified as cell boundary, cell interior or background. Such an approach can be very effective for scenarios where noise is not a big challenge, since it would be easy to distinguish between these three classes. However, as demonstrated in Figure 4, the definition of these classes is highly fluid in our dataset: It is possible to have both bright or dark edges and bright or dark cell interiors in the same patch, let alone the whole dataset. This makes it necessary to have a semantic segmentation approach rather than pixel-wise classification solution, in which the classes are well-defined.

3. Proposed Segmentation Model

We use a convolutional segmentation network, based on the SegNet[3] architecture. The original SegNet architecture consists of a convolutional encoder network followed by a decoder network. The encoder network is structurally similar to the widely used VGG16 network [15]. SegNet differs from its predecessors in that it does not need to learn how to upsample while decoding, since it keeps track of the indices chosen in the max-pooling operation during the encoding stage.

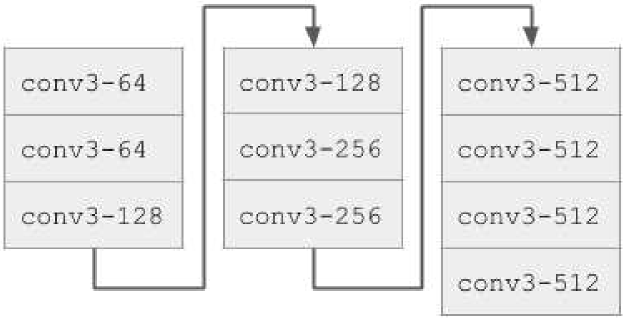


Figure 2: Encoder part of the network. In this figure, conv3 means a convolution layer with a receptive field of 3×3. Convolutional layers are followed by ReLU layers to introduce non-linearity. The decoder part, not shown here, is symmetric.

The structure of the encoder part of the network we used, which is based on the SegNet architecture is provided in Figure 2. In contract to the SegNet model, our model has only one channel in the input when only DIC is used. Additional channels can be added, depending on the experimental data. This encoder network is followed by a decoder network, which takes the output of the last layer of the encoder network and decodes it with successive deconvolution operations. The output at the end is a segmentation mask that quantifies the class scores for each pixel.

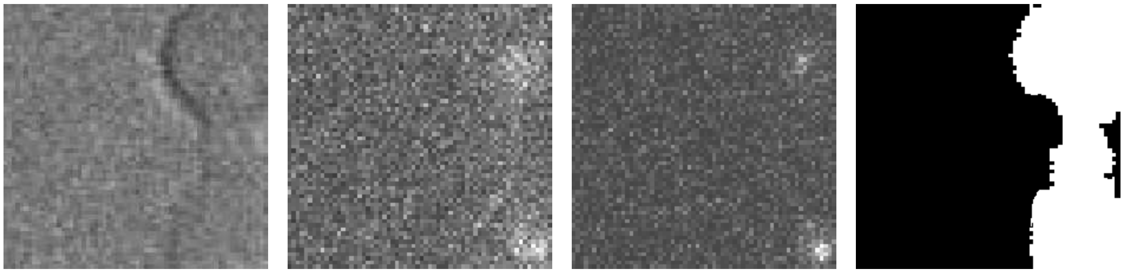
4. Dataset

We consider fluorescent microscopy images of the yeast cell division process. As mentioned in the introduction, these images are extracted from DNA-replication experiments, hence the illumination is purposefully limited, resulting in challenging conditions for segmentation. Our process for creating the dataset from the provided microscopy films is as follows: First, we extract an equidistant set of images from one of the experiments and manually label them using a custom web-based annotation tool (employing a superpixel algorithm). We randomly sample a set of patches of size 64×64 from each image. Since most of the image is empty background, we make sure that the cell region encompasses at least 10 percent of the patch, otherwise we pick another patch. After we select the patch, we apply standard data augmentation techniques (*i.e.,* flipping and rotating) to increase the sample size. Using this process we created a dataset of 6000 training samples, 1200 validation and 1200 test samples. Notice that training, validation and testing samples are extracted from independent experiments, which have varying numbers and positions of cells and degree of illumination. Demonstrating our results in such a dataset enables us to show that our model is generalizable.

Additional Channels. Beyond the DIC channel, which is the main channel in live-cell imaging, our dataset contains two more channels that are related to the particular experiment dynamics. Our goal in this paper is to show the possibility of using these secondary modalities to improve segmentation performance. In the live-cell experiment films that we obtained, multiple copies of a specific DNA sequence called *lacO* were inserted into a specific location in the yeast genome. Within the cell, these sequences are bound by a DNA-binding protein called Lac repressor, which is fused to a green fluorescent protein (GFP). The result of this process is that a specific site on a specific chromosome is labeled with a green fluorescent dot. This dot is represented as a bright spot in one of the channels, and lies with the rest of the DNA in the cell nucleus, which can be used as a marker for cell location. Similarly, multiple copies of another DNA sequence, tetO, are inserted into a nearby location in the genome, and are bound by the corresponding repressor tetR fused to a red fluorescent protein. Green dot intensity is observed through a green filter with a 488nm laser, while the red dot intensity has been observed through a red filter with a 561nm laser. The purpose of this experiment is to measure the increasing fluorescent intensity while the corresponding region on the chromosome is replicated, and this information is used to draw conclusions about the timing of DNA replication. A patch from our dataset with corresponding channels and ground truth mask is provided in Figure 3.

From the image segmentation perspective, having two more modalities in the dataset provide us more information about the location and the morphology of the cell. Since these channels display the luminescence of certain biological markers, bright regions correspond to the cell nucleus or cytoplasm, depending on the degree of brightness. One of the subsections in our experiments section, is devoted to the analysis of segmentation using these extra modalities.

Overall, the dataset we have created poses a challenging scenario for a cell segmentation algorithm. To illustrate this challenge, we annotated the same live-cell experiment frame twice with different human annotators. Then we sampled their annotations in the same way we created dataset (*i.e.,* cropped patches with more than 10 percent cell coverage) and found out that the agreement in terms of mIOU between the annotations of two human annotators is 0.79. We consider this to be a low agreement factor between the human annotators, which shows that the dataset is quite challenging, and provides us with a practical upper bound for the performance of an automated system.



(a) DIC channel (b) 488nm channel (c) 561nm channel (d) Ground truth mask

Figure 3: Different modalities in our dataset for a particular patch.

5. Results

5.1. Baseline Comparisons

We compared our architecture with several baselines and demonstrate that our method outperforms them in most settings. For the supervised models, we used the training, validation and test sets for their respective purposes. For the unsupervised techniques, we tuned the model parameters to come up with the best performance we can achieve on a single image, then measure the overall performance across the whole dataset with those model parameters. Results of our proposed method has been averaged over five runs. We used an Adam optimizer [8] to optimize the loss, with a learning rate of 2×10^{-3}, and with $\epsilon = 10^{-5}, \beta_1 = 0.9$ and $\beta_2 = 0.999$. We trained our model for 2 epochs with a batch size of 10.

Edge detection. This method starts with the detection of edges from the background using the image gradient. Detected edges are then dilated, and the region between them are filled to end up with a solid cell structure. The problem with this approach is that it requires hand tuning of the parameters for edge detection and dilation. Finding right set of parameters can be quite tedious, and may not be transferable across different experiments or even images.

Watershed transform. We implemented a watershed transform based segmentation method based on [16]. We created a hypothetical condition where we have information about the ratio of the overlap between a segment and the ground truth compared to the total area of that segment. We accepted segments that more than 40 percent of their total area falls over the ground truth. We created these conditions assuming such information may be available in the experiments we consider.

CRF. We trained a CRF (Conditional Random Field) model using the patches as train data, and tested with the test data. We used superpixel color, histogram of intensity values of the extracted patch, and HoG[7] features as the unary features, while we used color intensity difference, histogram difference and texture similarity[1] (using KL-divergence) as pairwise features.

CellProfiler. We used CellProfiler [6] to design a model that can accurately segment images. We used adaptive thresholding, since varying backgrounds due to high noise means a global threshold is likely to do more mistakes. We used background thresholding as our thresholding method, which utilizes intensity to decide whether a pixel belongs to foreground or background. The segmented image has been smoothed before the thresholding, using a Gaussian filter. We used a threshold correction factor of 0.7, < 1 being relatively lenient during the thresholding.

DeepCell We used the implementation provided by the authors of DeepCell[18] to train a pixel-wise classifier that can decide whether a pixel belongs to the background or cell interior based on a patch centered around the pixel. A variety of networks with different input sizes have been provided by the authors. We used the network with an input size of 61×61, since it is close to our patch size of 64×64 . For multi-modal data, the authors of DeepCell suggest the use of nuclear channel to predict the location for cell interior, followed by the prediction of cytoplasm location. While this approach is useful when there is a strong marker for nucleus, it is not suitable in our dataset since the nucleus marker channels themselves(*i.e.,* 488nm and 561nm channels) are quite noisy, and the regions corresponding to the cell nuclei become bright only in particular stages of the cell division. Hence, we followed the other pipeline suggested by the paper, which involves using only DIC channel, and thresholding the network's output. For thresholding, we found out that the adaptive thresholding methods fail, hence we manually chose the threshold that maximizes mIOU. Similarly, we removed connected components with size smaller than 400, which resulted in the highest mIOU. Since DeepCell produces output for the whole frame rather than the patches, we extracted the patches corresponding to our test dataset, and computed mIOU over these patches.

We compared the aforementioned techniques/algorithms to our proposed model. Our error metric is mean intersection over union (mIOU):

Method	**mIOU**
CellProfiler	0.3585
CRF	0.4261
Custom DeepCell	0.5644
Edge detection	0.5850
Watershed transform	0.6823
Proposed method	*0.7172*

Table 1: Comparison of mIOU in our dataset.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|P_i \cap G_i|}{|P_i \cup G_i|}, \tag{1}$$

where P_i is the set of predicted cell pixels and G_i is the ground-truth annotation for the cell pixels.

Results of our experiments are demonstrated in Table 1. Our method outperforms the baseline methods listed by a considerable margin. Considering the fact that the watershed transform has been provided with external information about the ground truth, we can assume that edge detection, followed by dilation region filling achieved the best performance among our baselines. Notice that DeepCell may not be suitable for the particular dataset we consider, since the datasets for which the authors suggest thresholding have a much cleaner background, while the nucleus markers used by DeepCell are usually much more visible and much less noisy, rendering both approaches relatively ineffective. It is possible that DeepCell can perform better than what we report here, although that may require post-processing with significant human involvement.

A sample of our segmentation results can be seen in Figure 4. Figure 4b is particularly important for us, as it provides us with examples of the cases where our method fails. As it can be seen in Figure 4b, many of the failure cases for our algorithm can be ameliorated using standard post-processing techniques. For example, the results for fifth and sixth column in Figure 4b can be improved by filling the empty(white) regions. However, there are still challenging cases as in the example in the 4th column, in which our algorithm fails to find a good segmentation, and it is not very likely to improve the result using a post-processing technique.

5.2. Multi-modal segmentation

Another task we wanted to accomplish in this paper is to see whether we can utilize the extra modalities of the data to achieve better segmentation results. To this end, we trained three separate segmentation networks using only DIC, only 488nm and only 561nm channels respectively. We then compared our predictions for the test dataset, and realized that in some cases, extra modalities are more effective as a basis for segmenting cells from the background. We found that the DIC channel achieves a better segmentation than other channels in roughly 71% of the test samples. However when 488nm and 561nm channels are used for segmentation 20% and 9% of the test samples are segmented better, respectively. mIOU score for the test data is 0.5136 for the network trained only with 488nm channel and 0.4263 for the network trained only with 561nm channel.

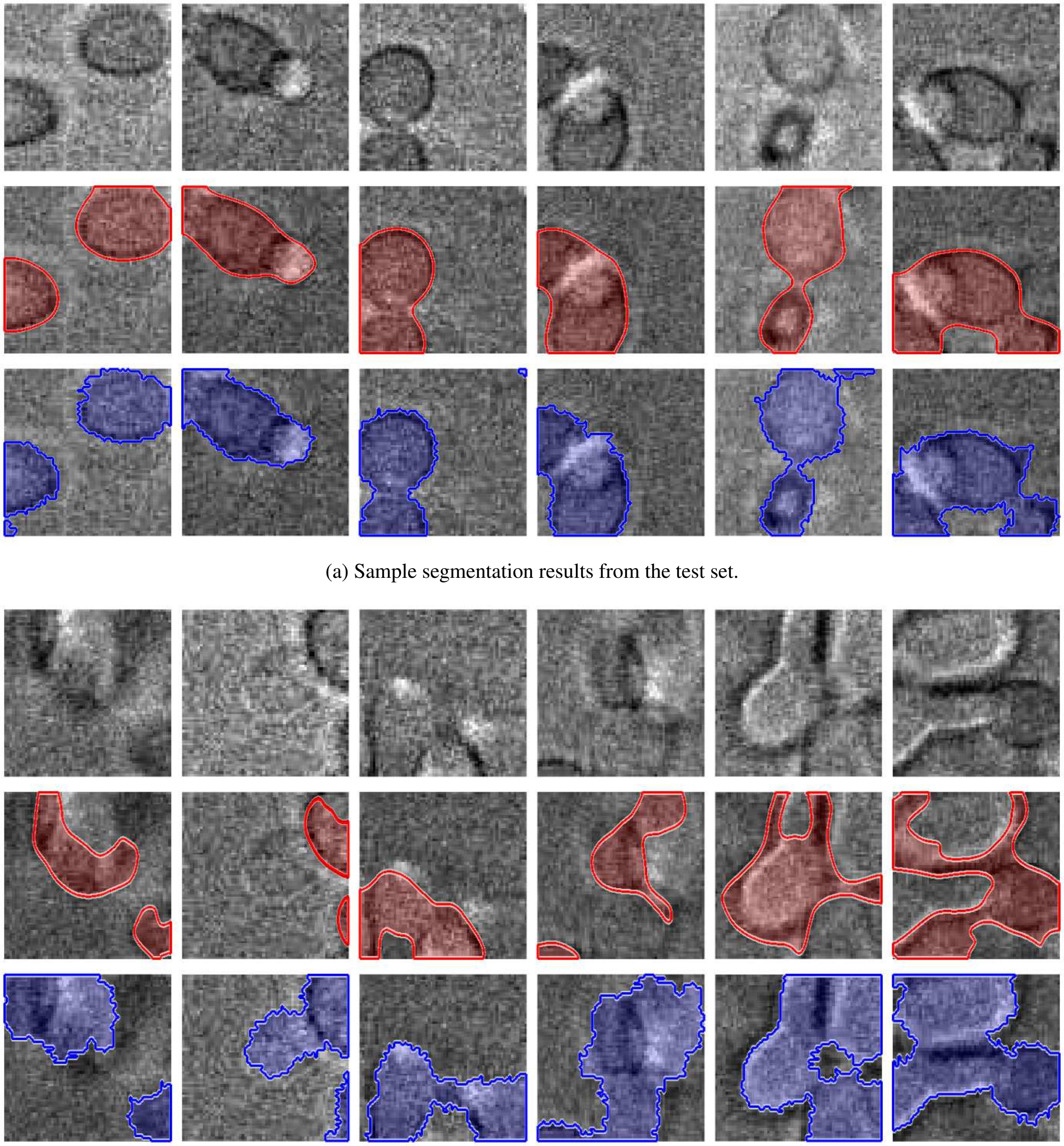
Method	**mIOU**
DIC only	0.7485
DIC + 488nm	0.7531
DIC + 561nm	0.7636
DIC + 488 nm + 561 nm	0.7709

Table 2: Comparison of training with different channels. Notice that these results are computed on a different dataset sampled from the same annotated data, hence not comparable with Table 1.

In order to measure the contribution of the extra channels, we sampled another dataset from our annotated data. Different from the initial dataset, our second dataset consisted of patches that are covered by cell regions at least 15% of their total area. We then compared the performance of our network trained only with DIC channel and the network fed with combinations of DIC and extra channels(DIC + 488nm, DIC + 561nm, DIC + 488nm + 561nm). Results of this comparison are provided in Table 2. Results in Table 2 shows a small yet easy-to-achieve improvement over the baseline, since extra channels are usually available for the experiments we are interested in and incorporating these channels into our model requires minimal effort. Extra channels can be fed to the network in the same fashion with feeding the individual channels for an RGB image, which is already possible in the SegNet architecture.

6. Conclusions

We argue that using a semantic segmentation network based on SegNet[3] can address many of the challenging cases while segmenting cells in live-cell experiments with low illumination. To exemplify this claim, we applied such a network to patches extracted from a live-cell experiment that involves the division process of yeast cells, and shot under low illumination to preserve experiment validity. The empirical results demonstrated that the network we utilized achieves a better result compared to many of the traditional methods that employ global parameters. Such methods usually fail because of the high variation in cell texture and the intensity of cell boundaries. Furthermore, we demonstrated that incorporating extra channels, which are often available in fluorescent microscopy experiments, into the segmentation network training process may improve the segmenta-



(a) Sample segmentation results from the test set.

(b) Sample cases from the test set where our method fails to generate a good segmentation.

Figure 4: In each image, top row corresponds to the original image (DIC channel), while middle row corresponds to the prediction and the bottom row corresponds to the ground truth.

tion performance, especially in the cases where the view obtained using DIC channel is not distinguishing enough to achieve a good segmentation.

7. Future Work

In Section 5.2, we demonstrated that the extra modalities can in some cases achieve higher segmentation performance, compared to the DIC channel. We plan to concentrate on this direction in our future work. Specifically, we plan to implement a semantic segmentation network that can fuse different modalities by correctly identifying which modality is more useful for a particular sample, and therefore select and output the segmentation mask for the best modality. Such a model is likely to achieve a significantly better overall segmentation performance compared to using only DIC channel, as hinted in Section 5.2. Extra modalities can be helpful specifically in challenging cases as the ones shown in Figure 4b, where cell regions are barely distinguishable from the background or from each other.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. *Computer vision-eccv 2004*, pages 469–481, 2004.

[2] D. Anoraganingrum. Cell segmentation with median filter and mathematical morphology operation. In *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pages 1043–1046. IEEE, 1999.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.

[4] P. Bamford and B. Lovell. Unsupervised cell nucleus segmentation with active contours. *Signal processing*, 71(2):203–213, 1998.

[5] K. Bredies and H. Wolinski. An active-contour based algorithm for the automated segmentation of dense yeast populations on transmission microscopy images. *Computing and Visualization in Science*, 14(7):341–352, 2011.

[6] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, et al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100, 2006.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[8] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] K. Z. Mao, P. Zhao, and P.-H. Tan. Supervised learning-based cell image segmentation for p53 immunohistochemistry. *IEEE Transactions on Biomedical Engineering*, 53(6):1153–1163, 2006.

[10] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5):530–549, 2004.

[11] E. Meijering. Cell segmentation: 50 years down the road [life sciences]. *IEEE Signal Processing Magazine*, 29(5):140–145, 2012.

[12] A. Paintdakhi, B. Parry, M. Campos, I. Irnov, J. Elf, I. Surovtsev, and C. Jacobs-Wagner. Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Molecular microbiology*, 99(4):767–777, 2016.

[13] M. E. Plissiti, C. Nikou, and A. Charchanti. Watershed-based segmentation of cell nuclei boundaries in pap smear images. In *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, pages 1–4. IEEE, 2010.

[14] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[16] I. The MathWorks. Marker-controlled watershed segmentation, 1999.

[17] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35):14271–14276, 2012.

[18] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Comput Biol*, 12(11):e1005177, 2016.

[19] C. Versari, S. Stoma, K. Batmanov, A. Llamosi, F. Mroz, A. Kaczmarek, M. Deyell, C. Lhoussaine, P. Hersen, and G. Batt. Long-term tracking of budding yeast cells in brightfield microscopy: Cellstar and the evaluation platform. *Journal of The Royal Society Interface*, 14(127):20160705, 2017.