Semantic segmentation of cell nuclei and cytoplasms in microscopy images

#[1] #
[1] *Address 1*
#[2] #
[2] *Address 2*
#[*3] #
[1]
#[*4] #
[1]

Editors: Under Review for MIDL 2019

Abstract

Microscopy imaging of cell nuclei and cytoplasms is a powerfull technique for research, diagnosis and drug discovery. However, the use of fluorescent microscopy imaging for cell nuclei and cytoplasms labeling is time consuming and inconvenient for several reasons, thus there is a lack of fast and accurate methods for prediction of fluorescence cell nuclei and cytoplasms from bright-field microscopy imaging. We present a method for labeling bright-field images using convolutional neural networks. We investigate different convolutional neural network architectures for cell nuclei and cytoplasms prediction. Using the DeepLabv3+, we found relative impressive results with a 5-fold cross validation dice coefficient equal to 0.9503 as well as meaningful segmentation maps. This work shows proof of concept regarding microscopy fluorescence labeling of cell nuclei and cytoplasms using bright-field images.

Keywords: Fully convolutional neural networks, semantic segmentation, deep learning, microscopy imaging, fluorescent imaging.

1. Introduction

Imaging of cell nuclei and cytoplasms is a powerful tool in diagnosis of diseases, research and drug development. Manual counting and estimation of quantitative measures of cells based on microscope images can be time consuming. A modern image cytometer, such as the # from #, facilitates cell analysis by automating the acquisition of the static microscopy images as well as the extraction of the quantitative characteristics of the cells using image analysis and statistics. In this modality, optical microscopy techniques are used to acquire static images of cells. In fluorescence microscopy imaging cells are stained with different fluorochromes yielding different label images. The fluorochromes in microscopy images are used to enhance contrast, and thus to infer labeling of cell nuclei and cytoplasms. In fluorescence microscopy the fluorochromes absorb incoming light and excites light with

* #

higher wavelength than the incoming light. Bright-field is the simplest label-free microscopy imaging technique, in which the contrast is related to the attenuation of white light. Bright-field microscopy imaging is a standard protocol, which is difficult to analyze, and manual identification of cell nuclei and cytoplasms is difficult. Therefore, the bright-field image is often acquired together with a fluorescence label of interest.

The use of fluorescent labelled images has several drawbacks; in certain circumstances the stain might effect the experiment, the stain might be toxic for the cells, fluorescent microscopy imaging requires preparation of the sample and finally spectral overlap gives physical limitations regarding simultaneous acquisition of multiple labels. Therefore, there is a high demand on fast and accurate methods for prediction of fluorescence labels from bright-field microscopy images.

Semantic segmentation is the process of assigning each pixel in an images into a label. Prediction of cell nuclei and cytoplasms from the bright-field image using the fluorescence labeled image as ground truth is thus a semantic segmentation task. The scope of this paper is to investigate methods for predicting pixels as nuclei or cytoplasms using convolutional neural networks. The input images for the convolutional neural network consist of bright-field images $\mathbf{x}$, and the output images $\mathbf{y}$ consist of processed fluorescence stained images representing the ground truth. These image are presented in Figure 2 together with prediction maps. The fluorescence image presented in Figure 2 is obtained using the fluorochromes DAPI (Kapuscinski, 1995) and #. This image is a categorical label image and is obtained by post-processing using # software as further described in section 3.

1.1. Related work

Several methods for cell segmentation based on convolutional neural networks (CNN) have been proposed in the recent years. Regarding cell segmentation on electron microscopy images, the so-called U-Net (Ronneberger et al., 2015) showed promising results on the ISBI 2015 cell tracking challenge using a fully convolutional neural network (FCNN). However, limited work has been published on cell nuclei and cytoplasms segmentation in bright-field imaging with fluorescence as the ground truth.

In (Christiansen et al., 2018) it is shown that it is possible to predict fluorescent labels from unlabeled microscopy images. In this work fluorescent labels are predicted using phase-contrast, bright-field and differential interference contrast imaging. Furthermore, multiple z-stack axis slices that are co-registered yielding 3D information content.

The models we investigate in this paper rely heavily on the prior work from the following. The backbone of the networks we experiment with is inspired by (Ronneberger et al., 2015), in which FCNN is utilized in an encoder-decoder scheme in combination with skip connections. The work by (He et al., 2016) utilizes residual blocks using identity mapping of input layers to circumvent vanishing gradient problems in order to train very deep convolutional neural networks. The alternative to the residual block, known as dense-block (Huang et al., 2017), uses concatenation of previous layers to improve the information flow in convolutional neural networks. The state-of-the art on cell segmentation relies on the backbone from U-Net with encoder decoder structure combined with some of blocks mentioned above. The work done by (Christiansen et al., 2018) is of certain interest since it deals with the same kind of data. However, there are certain differences: 1) the data provided from #

includes the whole cell specific stain BlueMask^{tm} from #, instead of CellMask^{tm}, which is a plasma membrane specific stain 2) the flourescent labeled images from # are processed before training using # software, which yields a categorical ground truth image consisting of background, cell nucleus and cell cytoplasm as described in Section 3, and finally 3) the architecture proposed by (Christiansen et al., 2018) uses the Inception module (Szegedy et al., 2015) and the residual block (He et al., 2016).

We investigate convolutional neural networks to predict post-processed fluorescence labeled cytoplasms and cell nuclei images from bright-field images using U-Net (Ronneberger et al., 2015), the Tiramisu (Jégou et al., 2017) and the DeepLabv3+ (Chen et al., 2018).

2. Methods

2.1. U-Net

We implemented a custom U-Net inspired by (Ronneberger et al., 2015). All convolutional layers use zero padding in order to retain spatial dimensions. The custom U-Net architecture consists of an encoding path with pooling and a decoding path with transpose convolution. Before the decoding path, a bottleneck is implemented with no pooling.
The architecture is illustrated in Figure 1. To prevent overfitting, l2 weight decay was utilized at each convolutional layer with w=0.0001.

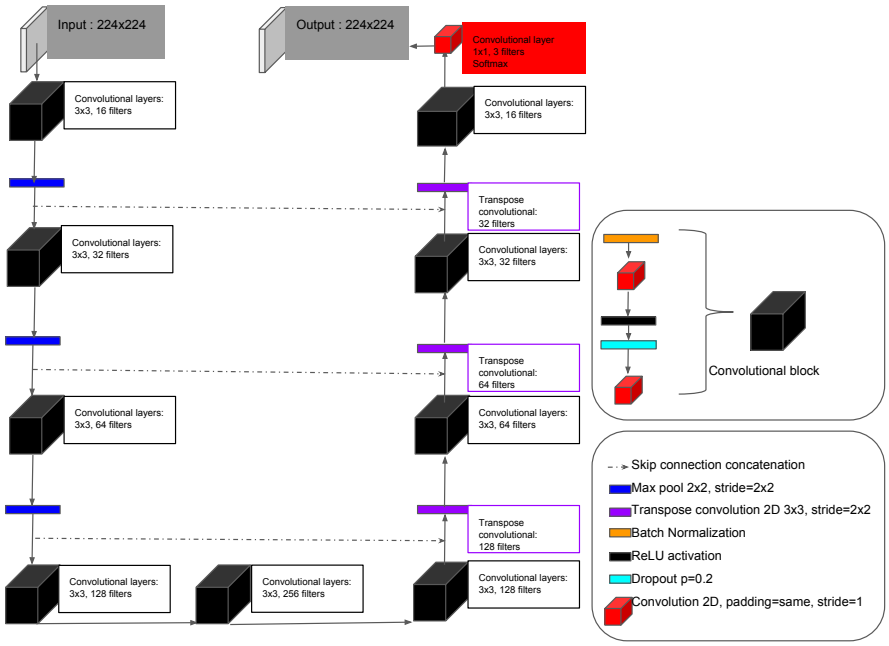

Figure 1: The network architecture inspired by U-Net

2.2. DeepLabv3+

The current state of the art for semantic segmentation of images on CityScapes (April 2018) and Pascal VOC (April 2018) is the DeepLabv3+ proposed in (Chen et al., 2018).

The DeepLab v3+ is a mixture of different components including; atrous convolution and the Xception block proposed in (Chollet, 2016), which consists of depthwise separable convolutions. The DeepLabv3+ utilizes also an encoder-decoder architecute like the U-Net. In the decoding path atrous spatial pyramid pooling (ASPP) is used for global image feature extraction.
Atrous convolution, also known as dilated convolution, is a powerful extension of ordinary convolution in which we can control the resolution of the receptive field of the convolutional operation (Yu and Koltun, 2015). Atrous convolution produces i spatial locations of the output feature maps $\mathbf{a}$ of the input $\mathbf{x}$ using the kernel $\mathbf{w}$.

$$\mathbf{a}[i] = \sum_k \mathbf{x}[i + r \cdot k]\mathbf{w}[k] \tag{1}$$

In this a hyperparameter atrous rate r describes rate in which one samples the input signal. The standard convolutional operater is thus a situation with $r = 1$. Atrous convolution is often implemented

Depthwise separable convolution is a combination of depthwise convolution and convolution with a 1x1 kernel (pointwise convolution). Depthwise convolution performs a convolution independently for each input channel. It is shown in (Howard et al., 2017) that depthwise separable convolution is significantly computational cheaper to calculate compared to standard convolution. In (Chollet, 2016), depthwise separable convolution is used, together with Residual blocks (He et al., 2016), as the building blocks to produce the Xception network, which outperformed the Inception network (Szegedy et al., 2015). The DeepLabv3+ architecture follows the original implementation in (Chen et al., 2018).

2.3. Tiramisu

Motivated by the work of (He et al., 2016) on ResNets, the authors in (Huang et al., 2017) develop the so-called DenseNets, which introduces the dense block. The dense block is an alternative to the residual block. The idea behind the dense block is to concatenate the identity mapping rather than adding it:

$$\boldsymbol{x_l} = H_l([\boldsymbol{x_0}, \boldsymbol{x_1}, ..., \boldsymbol{x_{l-1}}]) \tag{2}$$

In which $[\boldsymbol{x_0}, \boldsymbol{x_1}, ..., \boldsymbol{x_{l-1}}]$ refer to the skip connections which are concatenated from the previous layers l.

The composite function H_l refers to the cascade of operations including Batch Normalization, ReLU activation and 3x3 convolutions followed by dropout.

Inspired by the performance of DenseNets, (Jégou et al., 2017) introduced the Tiramisu network architecture for semantic segmentation. Along the encoding phase and the decoding phase, dense blocks are implemented. The encoding phase consist of dense blocks and max pooling (transition down) and the decoding path consist of transposed convolution (transition up). Before each dense-block in the downsampling path, features are forwarded and concatenated.

The Tiramisu architecture used in this paper follows the original implementation of the FC-DenseNet103 Tiramisu described in (Jégou et al., 2017).

3. Experiments

The input images are bright-field microscopy images and label images of adherent HeLa cells acquired with # device using 20x magnification. Since all the data originates from the same type of scanner the domain shift between the brightfield images is neglectable. The thresholding of fluorescence images to form ground-truth label images is achieved through unsupervised classification of pixels into the three classes; cytoplasm, nuclei, and background. In the dataset the cell nuclei appear app. 5 times more than the background, and the cytoplasms appear app. 10 times more than the background. From each sample two fluorescense images are acquired, one to detect nuclei pixels and one to detect cytoplasms pixels, respectively. The nuclei image is stained with DAPI and excited with wavelength equal to 405nm and detected at 593nm. The cytoplasms image is stained with # and excited at 405nm and detected at 452nm.

The provided data consist of 170 images with dimension: 1440x1920 together with corresponding label images of same size. The data was split into 153 training images and 17 validation images. Since the amount of data provided only consists of 170 images data augmentation was necessary in order to generalize and prevent over-fitting. The data augmentation includes; rotations of 90, 180, 270 degree, horizontal and vertical flip, and translation of 122 pixels vertical and horizontal. Furthermore, deformation using B-splines with 2 control points and standard deviation equal to 15 was utilized. For the DeepLabv3+, a 5-fold cross validation was carried out. The data was randomly split into 136 training images and 44 for validation.

Training a neural network with input size 1440x1920 is computational expensive and requires large amount of memory, but it is possible. However, training with large input size requires a smaller model. In order to train a large model and maintaining all possible information without resampling and to, patches of size 224x224 were sampled randomly from each category; background, cytoplasms and cell nuclei. Furthermore, this strategy of sampling from the different categories also compensates the class imbalance problem described previosly. All experiments were carried out on 4xTitan X GPU's. The training was done asynchronously on each GPU.

The implementation was carried out in Keras. We utilized a categorical cross entropy with the Adam optimizer (Kingma and Ba, 2014) with learning rate equal to 0.001, which drops by a factor of 0.5 every 50 epoch for 300 epochs. The mini-batch size was 3 pr GPU.

3.1. Results

The inspection of training and validation loss curves showed no sign of overfitting. The dice coefficient (Dice, 1945) is used to compare the found segmentation with the ground truth image. The performance of the different models are showed in Table 1. The DeepLabv3+ showed a validation dice coefficient at 0.9395. The bright field image, the ground truth flourescence label image and the prediction map is presented in Figure 2. A 5-fold cross validation for the DeepLabv3+ was carried out yielding a validation dice coefficient equal to 0.9503 and a validation loss equal to 0.14306.

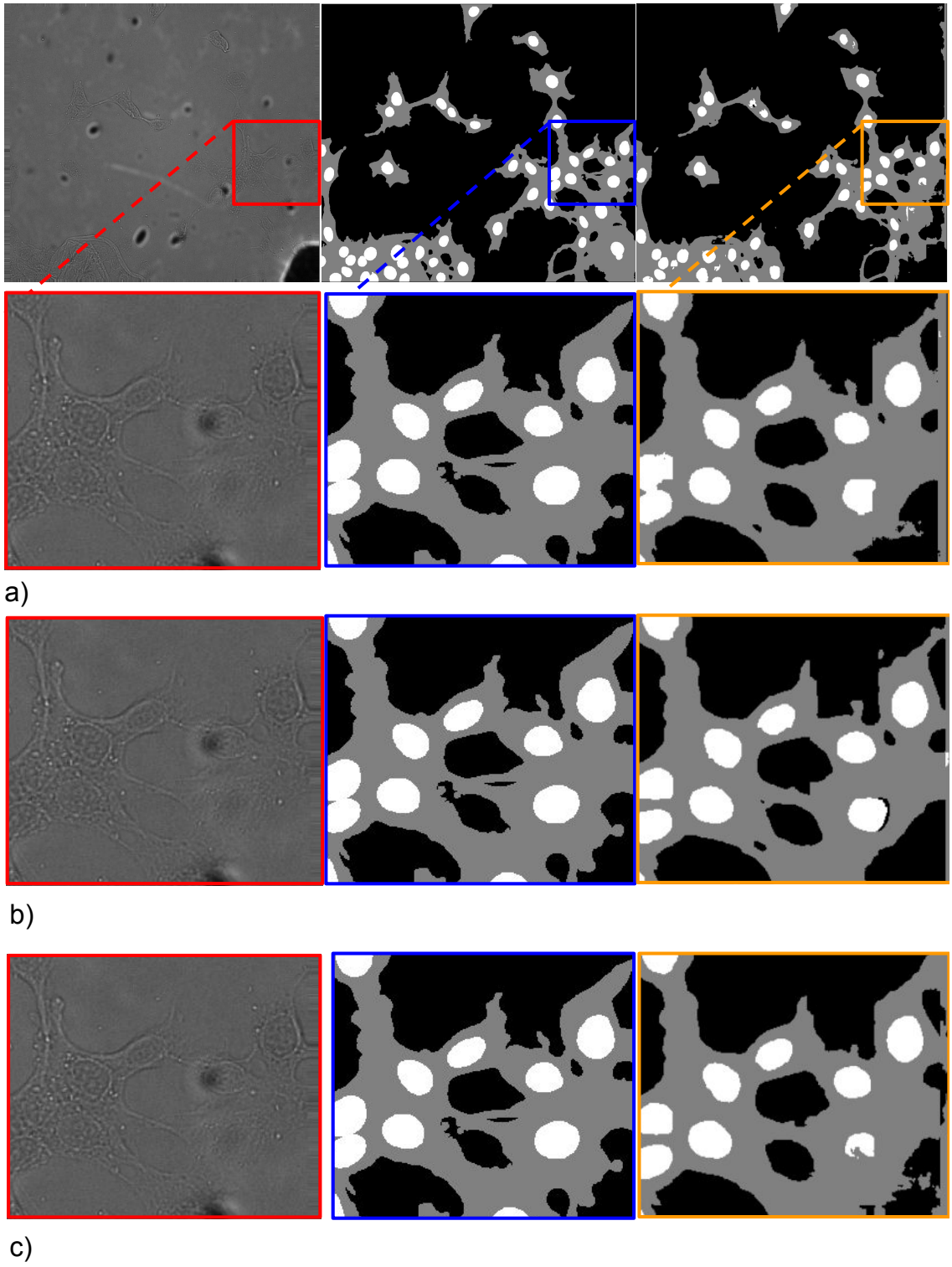

Figure 2: First row: Bright field image, flourescent label image and DeepLabv3+ prediction image. In second row to fourth row: Comparison between the three different models. a) The U-Net model, b) The Tiramisu model and c) the DeepLabv3+ model.

Table 1: Comparison of results

	U-Net	Tiramisu	Deeplabv3+	Deeplabv3+ (cross validation)
Parameters	1,942,023	9,419,011	41,253,023	41,253,023
Validation dice	0.9102	0.927	0.9395	0.9503

4. Discussion

As seen in the results, it is possible to predict fluorescent labels from bright-field images. The best performing architecture is the DeepLabv3+ as seen in Table 1. However, the performance gain with respect to the parameters using DeepLabv3+ compared to smaller models was not significantly large. This suggests that for the given segmentation task, a smaller model can be be sufficient to a certain extent. This makes sense since the domain of possible segmentations is limited to the same modality (bright-field), the same type of scanner and only three categories. Compared to (Christiansen et al., 2018), the proposed model has some common building blocks, however, it is impossible to compare the performance of the proposed model with the model in (Christiansen et al., 2018), since the data is not public available. However, we believe the proposed model shows state-of-the performance due to the high dice coefficient presented in Table 1.
Nevertheless, the model performance has certain limitations. In Figure 2 we see a big air bubble in the bottom right corner of the input bright-field image. Since cell nuclei and cytoplasms are not visible in this image, it is impossible for the model to correctly predict the labels. Future work will explore methods for mitigating problems with air bobbles and other difficulties in the data. Sophisticated pre-processing pipelines could be introduced to remove samples with air bobbles or other artifacts. Other input modalities can be acquired eg. phase-contrast of differential interference contrast, which can serve as input features and might improve the performance of the proposed models.

5. Conclusion

In this this work, we show proof of concept of cell nuclei and cytoplasms prediction from bright-field images with convincing results. A comparison of three state-of-the-art networks shows that the DeepLabv3+ is the best performing network but also the one with the highest number of parameters. Compared to other medical image segmentation tasks, we have a sufficient number of training samples making it possible to train the DeepLabv3+, which has a high number of parameters. The smaller networks also give usable results. We also show that the current limitations are due to noise in the acquisition that potentially could be resolved by adding information form other input modalities. Future work would reveal the impact of adding modalities like phase-contrast and differential interference contrast microscopy imaging. However, the results presented in this paper might be sufficient for research, drug discovery and medical diagnosis in which other sources of error might have greater impact.

References

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv preprint arXiv:1802.02611*, 2018.

François Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, 2016.

Eric M Christiansen, Samuel J Yang, D Michael Ando, Ashkan Javaherian, Gaia Skibinski, Scott Lipnick, Elliot Mount, Alison ONeil, Kevan Shah, Alicia K Lee, et al. In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell*, 2018.

Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26 (3):297–302, 1945.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, page 3, 2017.

Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1175–1183. IEEE, 2017.

Jan Kapuscinski. Dapi: a dna-specific fluorescent probe. *Biotechnic & Histochemistry*, 70 (5):220–233, 1995.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Olaf Ronneberger, Philipp Fischer, and Thomas 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.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, et al. Going deeper with convolutions. Cvpr, 2015.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.