Roman Bruch*, Rüdiger Rudolf, Ralf Mikut, and Markus Reischl

Evaluation of semi-supervised learning using sparse labeling to segment cell nuclei

https://doi.org/10.1515/cdbme-2020-3103

Abstract: The analysis of microscopic images from cell cultures plays an important role in the development of drugs. The segmentation of such images is a basic step to extract the viable information on which further evaluation steps are build. Classical image processing pipelines often fail under heterogeneous conditions. In the recent years deep neuronal networks gained attention due to their great potentials in image segmentation. One main pitfall of deep learning is often seen in the amount of labeled data required for training such models. Especially for 3D images the process to generate such data is tedious and time consuming and thus seen as a possible reason for the lack of establishment of deep learning models for 3D data. Efforts have been made to minimize the time needed to create labeled training data or to reduce the amount of labels needed for training. In this paper we present a new semi-supervised training method for image segmentation of microscopic cell recordings based on an iterative approach utilizing unlabeled data during training. This method helps to further reduce the amount of labels required to effectively train deep learning models for image segmentation. By labeling less than one percent of the training data, a performance of 90% compared to a full annotation with 342 nuclei can be achieved.

Keywords: Sparse labeling, Deep learning, Iterative training, Semi-supervised learning, Semantic segmentation

1 Introduction

Cell cultures can be used to examine the effectiveness and selectivity of an anti-cancer drug without the need to sacrifice animals. A large part of such studies relies on the evaluation of microscopic images, since they offer a wealth of information. From basic matters like the proliferation of cells up to more advanced aspects like the state of individual cells, a lot of questions can be answered on cell imaging.

***Corresponding author: Roman Bruch,** Institute of Molecular and Cell Biology, Faculty of Biotechnology, Mannheim University of Applied Sciences, Mannheim, Germany, e-mail: r.bruch@hs-mannheim.de
Rüdiger Rudolf, Institute of Molecular and Cell Biology, Faculty of Biotechnology, Mannheim University of Applied Sciences, Mannheim, Germany
Ralf Mikut, Markus Reischl, Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany

To extract this information in a quantitative and objective manner algorithms are needed. The segmentation of nuclei in microscopic images is a fundamental step on which further actions like cell counting or co-localization of other fluorescent markers depend on. Algorithms like built-in FIJI plugins [13], CellProfiler [9], Mathematica pipelines [14], TWANG [15] and XPIWIT [1] are well established for this task. Typically, these pipelines require properties such as object size and shape and therefore have to be reparameterized for different recording conditions or cell lines. In extreme cases, such as the segmentation of apoptotic cells, parameterization is not sufficient and special algorithms need to be designed [8].

In the recent years deep learning models like the U-Net [11] gained attention in the biological field due to their great modeling power: U-Nets can outperform classical segmentation methods [3], but therefore they need rich training data sets.

Newly emerging 3D cell cultures represent the living organism more closely than 2D cultures. Data sets are given as stacked image series. Furthermore, new difficulties such as decreasing brightness along the z-axis arise. Classical approaches, robust against intensity fluctuations exist, but suffer from parameterization [15]. Deep learning methods like the U-Net can also be used for 3D data [12], but they lack in establishment due to the effort to create 3D training data sets. The process of generating training data for 3D images is time consuming and burdensome: Since the visualization is effectively limited to 2D slices it is hard to conceive the object dimensions which often leads to a loss of overview. In a full manual approach, each plane in the 3D image has to be labeled individually. It is also challenging to achieve consistent segment boarders over consecutive planes [6].

As an example, to label a single $128x128x32px$ image patch containing 200 nuclei with the help of an interactive labeling method [16], a time of $7.5h$ was needed. To create a training data set with only ten image patches a time of $75h$ would be needed.

Thus, many methods were developed to reduce the 3D-labeling effort which can be divided in three major approaches: Interactive labeling [5, 16], weakly supervised learning [19] and artificial training data [2, 7, 17]. The goal of interactive labeling is to accelerate the annotation process by supporting the user in a semi-automatic manner. Weakly supervised learning uses different annotations like point or scribble an-

notations which are easier to generate. Artificial training data with known labels can be created from simulations or through unpaired image-to-image translations.

A relatively new approach called semi-supervised learning is emerging utilizing labeled and unlabeled data during training. Mittal et al. [10] proposed a two branch approach for semantic segmentation consisting of a GAN-based segmentation branch and a classification branch both utilizing labeled and unlabeled data during training. Xie et al. [18] utilized a student teacher approach to incorporate unlabeled images in their training routine.

In this paper we introduce a new method for semi-supervised learning which utilizes iterative training in combination with a new approach to extract labels based on the three class approach used for sparse labeling in [6]. The potential of the introduced method is analyzed on 2D microscopic recordings of cell nuclei. We compare and investigate the influence of the number of initial labels in relation to a fully labeled data set, evaluate the progress over successive iterations and show typical errors that occur during training.

2 Methods and Data

The deep learning model used for this method is a modified U-Net [12] with a filter size of 32 in the first block, since it is proven that U-Nets work with sparsely annotated data [6]. The model is trained to minimize a weighted binary cross entropy to avoid negative effects resulting from class imbalances.

Figure 1 shows the concept of the iterative training. For the initialization of the method a sparsely annotated data set is required. Since the goal is to minimize the labeling effort, the user annotates only a few objects representative for the data set including also their surrounding background. To further decrease the annotation effort, interactive labeling methods can be used for this in the future.

The resulting labels then contain three classes: "ignored", "background" and "foreground". The ignored class indicates pixels for which no labels are provided and which training weights are therefore set to zero [6].

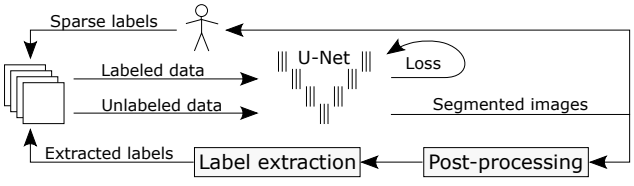

Fig. 1: Concept of the iterative training. First, sparsely labeled data is used to train a U-Net which is then used to segment the unlabeled data. After a post-processing step new labels are extracted from the segmented image which are then used in the next training step.

After training, the U-net is used to segment the unlabeled data and a post-processing step consisting of a morphological closing using a 3×3 square structuring element is performed to remove noise and to smooth the boundaries. Afterwards each segmented object is analyzed and if the mean confidence of the raw prediction is greater than a threshold $t = 0.8$, it is added to the training data or otherwise discarded and thus labeled as ignored. All of the segmented background is assigned to the label class background, regardless of confidence. In contrast to labels from previous iterations, the initial user defined labels will not be overwritten. The training data for the next iteration thus contains the user labels and the newly extracted labels. A new U-Net is then trained with the updated training data and new labels can be extracted to continue the loop.

During training, the data is randomly augmented including random noise, scale and contrast to enable the network to segment objects slightly different from the labeled ones [18]. To avoid overfitting to the training data, 25% of the objects are withheld from training to serve as validation data. The training of each iteration is continued until the performance on the validation set is not further increasing.

To evaluate the performance of the iterative training method for varying amounts of initially labeled nuclei and differing areas of nearby included background a subset of the publicly available data set BBBC039v1 [4] is used. Gaussian noise with $\sigma = 0.01$ was applied to the images to slightly increase the difficulty. The data set used is small, but sufficient to discuss effects based on a few representative examples.

The training was performed on three images shown in Figure 2 in which different amounts of nuclei ranging from 1 to 100 are labeled. The nuclei where randomly selected for each experiment. The nearby background of the selected nuclei is also labeled. Two sizes defined by 10 and 20 successive morphological dilation operations with a 3×3 square structuring element were tested. These two sizes get referred to narrow and broad background in the following. For validation, a separate image with a 25% lower number of labeled nuclei is used.

The performance of the trained models is measured using the average Intersection over Union (IoU) of ten repetitions on a test set consisting of 15 completely labeled images. The IoU score was normalized to the mean IoU achieved on the test set by training with a completely labeled train and validation set.

3 Results

By labeling $3, 30$ and 100 nuclei including the broad background a normalized IoU of $90.2\%, 97\%$ and 100% compared

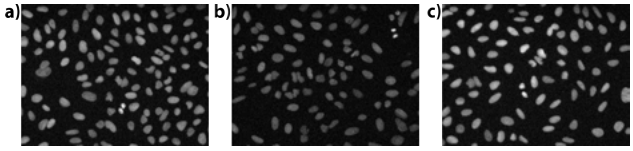

Fig. 2: Three images used for training of the models.

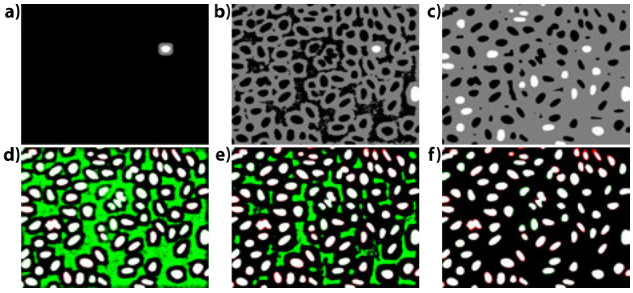

Fig. 3: Results with one initially labeled nucleus and narrow background. It shows the initial/extracted labels (a-c) and binarized predictions (d-f) of the iteration 0, 1 and 3. (a-c): black, gray and white pixel represent the classes ignored, background and foreground. (d-f): green and red indicate false positive and false negative pixel errors. The original image can be seen in Figure 2 (c).

to a full annotation with 342 nuclei is achieved after three training steps. The performance for the narrow background is worse as only a normalized IoU of $77.8\%, 81\%$ and 79.1% is obtained.

Figure 3 shows results for a training with a single labeled nucleus and narrow background. While the segmentation after the initial training with one labeled nucleus suffers from large areas with wrongly segmented background (d), the next models can recover and correctly segment these areas after the fourth iteration (f). The image also shows that errors can aggregate over subsequent iterations, as is the case for the elongated nucleus at the top right of the image, where the segmentation is shrinking.

To see whether this trend is common across all numbers of initial labeled nuclei, the mean percentage of false positive and false negative pixels on the training images from ten repetitions was measured over the iterations and visualized in Figure 4.

As can be seen, large quantities of false positive pixels are common to all amounts of labeled nuclei, whereas it can be greatly reduced over subsequent iterations. With the broad background the amount of false positive pixels can be reduced and an increase in the amount of labeled nuclei can further decrease the error. The figure also shows that the amount of false positive pixels is slightly increasing over the iterations. Labeling of only few nuclei reduces the amount of false positives, but increases the amount of false negatives.

4 Discussion

The results show that the proposed method can increase the performance on a sparsely labeled data set. By labeling less than 1% of the training data (3 nuclei), a performance of 90% can be achieved compared to a full annotation with 342 nuclei.

Large background areas in which no nuclei are located seem to be a problem of sparse labeling when only regions nearby nuclei are labeled. Since these areas can make up a large part of the images, results are greatly influenced by them. The problem can be reduced by annotating a greater region around the labeled nuclei (broad background). Specific annotation of few of such regions could possibly solve this problem completely. Nonetheless the iterative training method can recover from such mistakes and thus greatly improve the segmentation performance. This can be beneficial for 3D data as it is hard to label big volumes of background regions due to the limitation of 2D visualization.

The results show that the method can reduce the number of false positive pixels, while also leading to a slow increase of false negative pixels over subsequent iterations. The main reason for this effect seems to be the vanishing nuclei. Misclassified parts of a nucleus are used as a background labels in the next iteration, which results in a further increase of false negative pixels for that nucleus.

The enlargement of ignored regions to avoid the use of false negative pixels as labels did not improve the situation as it leads to an increase in size of all segmented nuclei and thus to an increase of false positive pixels. Attempts can be made to specifically enlarge only the ignored regions for nuclei which do show a low confidence. Moreover an advanced post-processing strategy can be implemented, like an active contour method to revise the segmentation. We also can think of a user intervention after a few iterations to select good segmented nuclei.

At the current state the method seems to be most useful when only labeling a few nuclei including as much background as can be quickly annotated. An optimal stop-criterion is still to be found, but a possible criterion could be vanishing changes in the percentage of ignored labeled pixels.

This workflow will also be adapted for 3D images in future work. There the user would sparsely label 2D slices of different orientations or complete nuclei including their nearby background while the extraction of labels can be performed in the same way just on 2D slices. An extraction based on volumes may also be used, but the confidence of complete nuclei could be too low.

Author Statement

Research funding: RR was funded by a grant from the Carl-Zeiss Stiftung. This work was funded by the German Federal

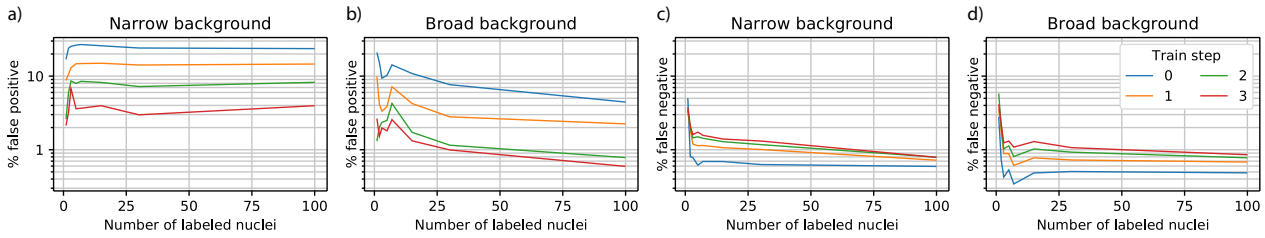

Fig. 4: Percentage of false positive (a,b) and false negative (c,d) pixels on the training set for three train steps and different amounts of labeled nuclei including the narrow (a,c) or broad (b,d) background. The percentage values are shown on a logarithmic scale.

Ministry of Education and Research (BMBF) as part of the Innovation Partnership M2Aind, project M2OGA (03FH8I02IA) within the framework Starke Fachhochschulen–Impuls für die Region (FH-Impuls). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The conducted research is not related to either human or animal use.

References

[1] Bartschat A, Hübner E, Reischl M, Mikut R, Stegmaier J. XPIWIT—an XML pipeline wrapper for the insight toolkit. *Bioinformatics* 2015;32:315–317.

[2] Böhland M, Scherr T, Bartschat A, Mikut R, Reischl M. Influence of synthetic label image object properties on GAN supported segmentation pipelines. In: Proc., 29. Workshop Computational Intelligence, Dortmund. 2019, 289–309.

[3] Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, et al. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature Methods* 2019;16:1247–1253.

[4] Caicedo JC, Roth J, Goodman A, Becker T, Karhohs KW, Broisin M, et al. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry Part A* 2019;95:952–965.

[5] Chen J, Ding L, Viana MP, Hendershott MC, Yang R, Mueller IA, et al. The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images. *bioRxiv* 2018;:491035.

[6] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D u-net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI Springer International Publishing 2016, 424–432.

[7] Dunn KW, Fu C, Ho DJ, Lee S, Han S, Salama P, et al. DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Scientific Reports* 2019;9.

[8] Khan AuM, Weiss C, Schweitzer B, Hansjosten I, Mikut R, Reischl M. Multimodal image segmentation of cellular fragmentation using edge detector and morphological operators. *Biomedizinische Technik/Biomedical Engineering* 2014; 59:518–521.

[9] McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biology* 2018;16:1–17.

[10] Mittal S, Tatarchenko M, Brox T. Semi-supervised semantic segmentation with high- and low-level consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2019;:1–1.

[11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science Springer International Publishing 2015, 234–241.

[12] Scherr T, Bartschat A, Reischl M, Stegmaier J, Mikut R. Best practices in deep learning-based segmentation of microscopy images. In: Proc., 28. Workshop Computational Intelligence, Dortmund. 2018, 175–195.

[13] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 2012;9:676–682.

[14] Schmitz A, Fischer SC, Mattheyer C, Pampaloni F, Stelzer EHK. Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids. *Scientific Reports* 2017;7:1–13.

[15] Stegmaier J, Otte JC, Kobitski A, Bartschat A, Garcia A, Nienhaus GU, et al. Fast segmentation of stained nuclei in terabyte-scale, time resolved 3d microscopy image stacks. *PLoS ONE* 2014;9:e90036.

[16] Tasnadi EA, Toth T, Kovacs M, Diosdi A, Pampaloni F, Molnar J, et al. 3D-cell-annotator: an open-source active surface tool for single cell segmentation in 3D microscopy images. *bioRxiv* 2019;:677294.

[17] Wiesner D, Svoboda D, Maška M, Kozubek M. CytoPacq: a web-interface for simulating multi-dimensional cell imaging. *Bioinformatics* 2019;35:4531–4533.

[18] Xie Q, Hovy E, Luong MT, Le QV. Self-training with noisy student improves ImageNet classification. *arXiv preprint arXiv:191104252* 2019.

[19] Yoo I, Yoo D, Paeng K. Pseudoedgenet: Nuclei segmentation only with point annotations. *arXiv preprint arXiv:190602924* 2019.