Search for temporal cell segmentation robustness in phase-contrast microscopy videos

Estibaliz Gómez-de-Mariscal[1]
[1] *Bioeng. and Aerospace Eng. Depar., Universidad Carlos III de Madrid, 28911 Leganés, Spain*
Hasini Jayatilaka[2]
[3] *AtlasXomics Inc. New Haven, CT, 06511, USA*
Özgün Çiçek[3]
[4] *Department of Computer Science, Albert-Ludwigs-University, Freiburg 79110, Germany*
Thomas Brox[3]
Denis Wirtz [4,5]
[5] *Dep. of Chemical and Biomol. Eng., The Johns Hopkins University, Baltimore, MA 21218, USA*
[6] *Dep. of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MA 21205, USA*
Arrate Muñoz-Barrutia[1] MAMUNOZ@ING.UC3M.ES

Editors: Under Review for MIDL 2022

Abstract

This work presents a deep learning-based workflow to segment cancer cells embedded in 3D collagen matrices and imaged with phase-contrast microscopy under low magnification and strong background noise conditions. Due to the experimental and imaging set-up, cell and protrusion appearance change largely from frame to frame. We use transfer learning and recurrent convolutional long-short term memory units to exploit the temporal information, and provide temporally stable results. Our results show that the proposed approach is robust to weight initialization and training data sampling.

Keywords: Cell segmentation, transfer-learning, convlstm, phase-contrast microscopy

1. Introduction

Studying cell morphology changes in time is critical to understanding cell migration mechanisms. In this work, we present a deep learning-based workflow to segment cancer cells embedded in 3D collagen matrices (more physiologically relevant that 2D cell cultures) and imaged with phase-contrast microscopy (on a fixed focal plane) at low resolution. Due to the experimental and imaging set-up, the images are contaminated with a strong noise and a high presence of artifacts. Moreover, there is a high variability in cell body and protrusions appearance from frame to frame. Previous contributions to the problem of cell segmentation on phase-contrast microscopy images (Arbelle and Raviv, 2019) mainly focus on improving the convolutional neural network (CNN) architecture and, sometimes, the loss function rather than the training strategy. However, the use of pre-trained encoders, the approach to deal with data imbalance, or to avoid the creation of artifacts during data augmentation are also important aspects to consider. Additionally, here we introduce temporal-consistency as we needed it to achieve a robust segmentation of cells exiting and entering the plane of focus, fixed along with the videos. Inspired by the work of Arbelle *et al.* (Arbelle and Raviv, 2019), we propose to use a recurrent architecture to segment cells in our low quality phase contrast microscopy videos. For this, we build the ground truth dataset from the

videos acquired for (Jayatilaka et al., 2018). Human fibrosarcoma HT1080WT (ATCC) cells were embedded in 3D collagen type I matrices and imaged with a 10X magnification every 2 minutes for 16.7 hours. The images covered a field of view of 1002 $\times$ 1004 pixels (pixel size of 0.802$\mu m/pixel$). The videos were pre-processed for drift artifacts, resulting in a final size of 983 $\times$ 985 pixels. We subtract 56 short sections between 3 and 100 frames from 27 videos corresponding to independent replicates, resulting in a total of 992 frames. Data available at https://zenodo.org/record/5979761.

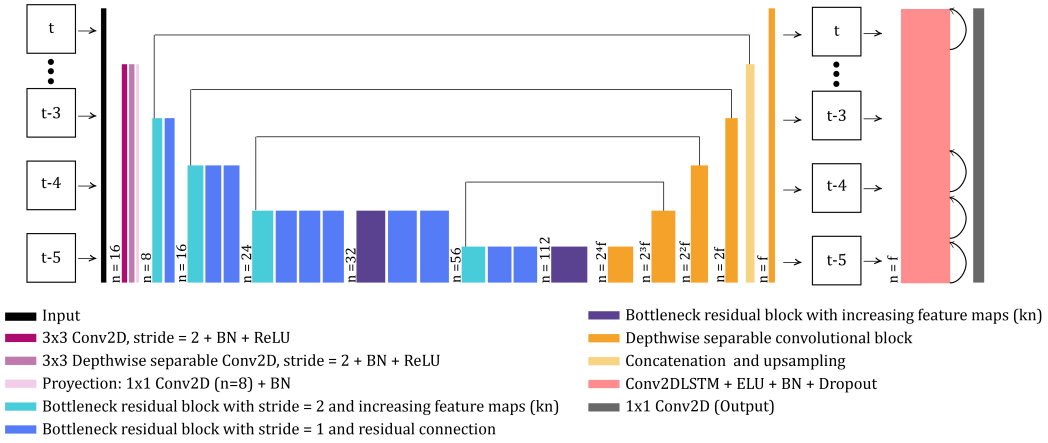

Figure 1: Convolutional neural network (CNN) architecture exploits spatio-temporal information.

	Pools/Conv. filters/Batch	TF	FT
B1	5/16-32-64-128-256/1	0.473	0.430
B2	4/25-50-100-200/1	0.411	0.455
B3	4/25-50-100-200/2	0.550	0.551
B4	4/50-100-200-400/1	0.481	0.446
B5	3/16-32-64	2/0.437	0.561

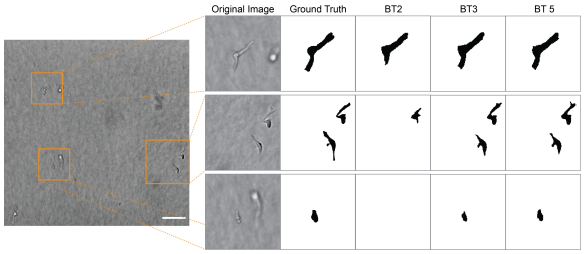

Figure 2: (Left) Proposed architecture and SEG for transfer learning (TF) and fine tuning (FT). The size of the convolutional layers only applies for the decoder. (Center) Original phase-contrast microscopy crop. (Right) Ground Truth; Output of the trained models B2, B3 and B5. Scale bar of 100 μm.

2. Methodology

We propose building a recurrent U-Net that uses (1) a pre-trained encoder (35% of the convolutional filters of the MobileNetV2 pre-trained on the ImageNet) with skip connections to the decoder, (2) a single ConvLSTM layer at the end of the encoder-decoder to optimize the memory usage, (3) one last 2D convolutional layer with two channels (background and foreground), and (4) separable depth-wise convolutions. The last recurrent layer will sequentially analyze the frames of an input video once the 2D U-Net processes them. This architecture provides segmentation for the frame at time t based on the information in the frames $(t-k, ..., t)$ for a chosen time window k. See Figure 1. We propose a training

data sampling strategy that first, transforms the image and second, crops a patch using a probability distribution function that deals with the unbalanced foreground-background ratio on the fly. The implementation and further details are available at https://github.com/esgomezm/microscopy-dl-suite-tf.

3. Experimental results

A set of 28 patches from the training dataset were used as a validation set. Each input image intensity is normalized using the percentile normalization (0.1 and 99.1). First, we used transfer-learning of the MobileNetV2 weights with a learning rate of 0.005, by freezing the encoder and randomly initializing the decoder. Second, we unfreeze the encoder and fine-tune all the weights using a learning rate of 0.0001 with a reduction schedule. Five different configurations of the proposed architecture Bi, $i = 1, ..., 5$ are tested. See Figure 1. During the training, we evaluate the binary segmentation with the Jaccard index on the validation set and then, we employed the Evaluation Software provided in the Cell Tracking Challenge (http://celltrackingchallenge.net/, SEG measure) to assess the results with the test set.

4. Discussion and conclusion

Overall, the proposed approach improved the robustness to intra-cell variations along time (i.e., caused by cell exiting and entering the plane of focus), achieving more consistent results at the cell edges. A critical point in our problem is the trade-off between image resolution and network configuration. The thinnest details, i.e., the cell protrusions, are the primary source of error for the segmentation. Low resolution hinders the learning of essential features and voids the effect that those pixels may have in the loss function or gradients. A potential computational experiment could be to upsample the size of the images and repeat the training process. B3 and B5 are the most accurate CNNs, coinciding with those cases trained with a larger batch size (Figure 1). We tested the impact of using temporal information by training a model with the same architecture as B5 but with only one input image. It showed that integrating the temporal information improved the segmentation result. The combination of whole cell shape segmentation and tracking will be already a promising tool for studying new clinical strategies to mitigate the metastatic phenotype.

Acknowledgments

This work was co-financed by ERDF, "A way of making Europe" (AMB), partially funded under Grant PID2019-109820RB-I00, MCIN/AEI/10.13039/501100011033/; the US NIH under Grants UO1AG060903 (DW) and U54CA143868 (DW). We acknowledge NVIDIA Corporation for the donation of the Titan X (Pascal) GPU.

References

Assaf Arbelle and Tammy Riklin Raviv. Microscopy Cell Segmentation Via Convolutional LSTM Networks. In *2019 IEEE 16th Int. Symposium on Biomedical Imaging (ISBI 2019)*, 2019.

Hasini Jayatilaka, Anjil Giri, Michelle Karl, Ivie Aifuwa, Nicholaus J Trenton, Jude M Phillip, Shyam Khatau, and Denis Wirtz. EB1 and cytoplasmic dynein mediate protrusion dynamics for efficient 3-dimensional cell migration. *FASEB J.*, 32(3):1207–1221, 2018.