Enhancement of 2D U-Net biomedical segmentation model with ImageNet pretrained CNNs

Maciej Szymkowski[1] M.SZYMKOWSKI@PB.EDU.PL
[1] *Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland*

Editors: Under Review for MIDL 2025

Abstract

Segmentation is one of the most important steps in medical and biomedical studies. It allows retrieval of the object and removal of unnecessary background. In this work, we propose the enhancement of 2D U-Net with ImageNet-pre-trained models (ResNet-50, VGG-16 and InceptionResNetV2) used as backbones. Part of the layers of these models were transferred to the 2D U-Net encoder (one CNN layers per experiment). The modified segmentation model was trained, tested, and evaluated on the set of frames from spatial video showing a single Engineered Heart Tissue (EHT) cell. The data set consisted of 1240, 140, and 100 frames for training, testing and evaluation, respectively. Two metrics were calculated during evaluation - in the best case Dice reached 0.945705 whilst Intersection over Union (IoU) was equal to 0.897424.

Keywords: 2D U-Net, Engineered Heart Tissue (EHT), ResNet-50, VGG-16, InceptionResNetV2

1. Introduction

Organ on Chip (OOC) is a microfluidic device that allows simulation of the behavior of specific organ's tissues or even single cells. It's main aim is to reduce or completely eliminate testing of new medications on animals. However, despite all advantages, that solution has one major disadvantage - the change in the tissue or cell behavior may occur even after several hours from substance application. It can be tiring to observe the object during that time (even with specialized tools). It is also a reason why biologists record the videos of the cell - they can then analyze the behavior frame by frame. To reduce the complexity of that method, we would like to propose fully-automatic model for the tissue/cell behavior analysis. It's first stage is segmentation - we would like to analyze the object without unnecessary background.

One of the well-known models used for biomedical segmentation is U-Net (Ronneberger et al., 2015). There are several works in which this architecture is modified to increase its segmentation ability, for example, U-NeXt (Valanarasu and Patel, 2022), U-Net++ (Zhou et al., 2018), or nnU-Net (Isensee et al., 2018). On the other hand, there are also papers that try to combine the U-Net architecture with the idea of Vision Transformer (Dosovitskiy et al., 2021). Significant examples of these works are Trans-U-Net (Chen et al., 2021) or DSTrans-U-Net (Lin et al., 2021). On the other hand, there are also works were Neural Cellular Automata (NCA) is used for segmentation of medical samples (Kalkhof et al., 2023). However, it must be claimed that in neither of these works transfer of selected layers from pre-trained models is performed to improve the model outcome.

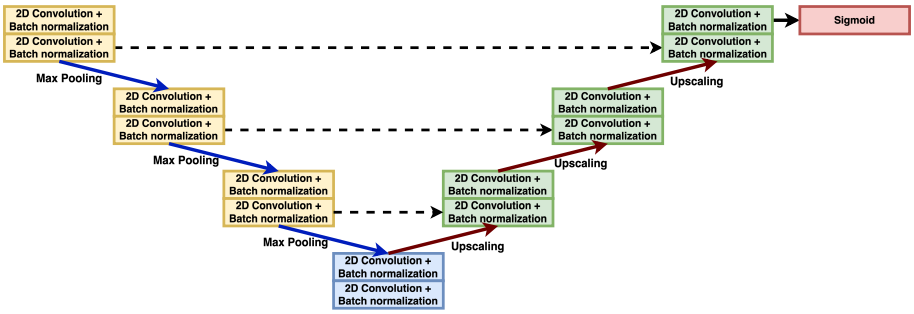


Figure 1: 2D U-Net scheme used for segmentation of the cell

In the scope of this work, we introduce an idea of usage of selected layers from previously trained (with ImageNet) convolutional neural networks (CNNs) in 2D U-Net encoder (the decoder is not directly affected by any layer of CNN). We consumed three pre-trained models - ResNet-50 (He et al., 2015), VGG-16 (Simonyan and Zisserman, 2015), and InceptionResNetV2 (Szegedy et al., 2016). The goal of their usage was to check whether the improvement of single EHT cell segmentation is observable. In the further part of the work, we present the results of the performed experiments.

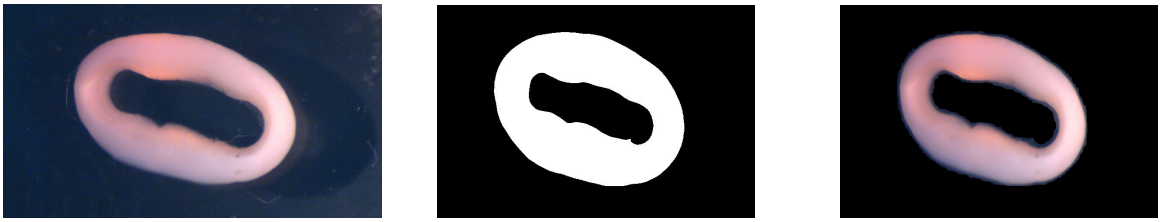
2. Proposed approach and experiments results

The approach proposed in this research is based on 2D U-Net segmentation model, presented in Fig. 1. At the very beginning, we prepared two strategies to tackle the problem of biomedical images segmentation. The first was to train the model from scratch - it means that the weights were initialized randomly (with Xavier initialization). In the scope of the second approach, we proposed usage of selected layers from CNNs pretrained on ImageNet (ResNet-50, VGG-16, and InceptionResNetV2 - during training, we consumed the layers from one model at a time, layers from different CNNs were not mixed) in U-Net encoder (decoder was not affected and it was randomly initialized). It was a kind of "knowledge transfer" from CNN to U-Net segmentation model. During experimental phase, two loss functions were used - Focal loss (FL) and Binary Cross Entropy (BCE). It must be also pointed out that the first scheme (without prior knowledge) was trained within 40 epochs whilst the training of the models with CNN backbone lasted 20 epochs (their convergence was much faster due to the prior knowledge). Learning rate in both cases was equal 0.001. The results of the performed trainings are given in Table 1. Example of the segmentation results - mask and final image are shown in Fig. 2.

It is also worth mentioning that training was performed with 1240 images while testing and evaluation with 140 and 100 samples respectively (all of them were collected by the Team of prof. Tomasz Kolanowski from Institute of Human Genetics, Polish Academy of Science). Moreover, it was not only observed that usage of CNN backbone (in 2D U-Net encoder) can shorten the training but also that it leads to the higher model metrics - both Dice and IoU were lower in the case of 2D U-Net trained from scratch. One more important conclusion is that the model with backbone was more certain about the segmentation results - when it comes to training from scratch, pixels with probability of 40% of belonging to the

Table 1: Results of the performed experiments - Dice and IoU calculated for different models

Network	**Backbone**	**Freeze**	**Dice**	**IoU**
U-Net with BCE	N/A	N/A	0.301895	0.183314
U-Net with FL ($\alpha = 0.8$)	N/A	N/A	0.915201	0.848639
U-Net with FL ($\alpha = 0.8$)	ResNet-50	YES	0.945546	0.897186
U-Net with BCE	ResNet-50	YES	0.933888	0.876329
U-Net with FL ($\alpha = 0.8$)	**ResNet-50**	**NO**	**0.945705**	**0.897424**
U-Net with FL ($\alpha = 0.8$)	VGG-16	YES	0.939017	0.885833
U-Net with BCE	VGG-16	YES	0.937961	0.883814
U-Net with FL ($\alpha = 0.8$)	VGG-16	NO	0.943959	0.894383
U-Net with FL ($\alpha = 0.8$)	InceptionResNetV2	YES	0.937819	0.883294
U-Net with BCE	InceptionResNetV2	YES	0.941037	0.889093
U-Net with FL ($\alpha = 0.8$)	InceptionResNetV2	NO	0.943449	0.893321



(a) (b) (c)

Figure 2: Segmentation process: original image (a), mask received from model (b) and segmented sample (c)

object were assigned as a part of its structure whilst this threshold was equal 90% in the case of models with CNN backbone.

3. Conclusions and future work

The main goal of these experiments was to check whether usage of different backbones in the 2D U-Net encoder can have an influence on biomedical samples segmentation quality. The outcomes of the performed research lead to the conclusion that specific backbones can improve the model. Both Dice and IoU were increased (in the best case the differences were 0.0305 and 0.0488 for Dice and IoU respectively) whilst also changes in quality of segmented images were visible during visual inspection of different models outcomes. Right now, we are working on enlarging the database, in the next stages we would like to repeat the performed experiments with at least 20-30k samples to confirm the obtained results with much broader datasets.

References

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation, 2021. URL https://arxiv.org/abs/2102.04306.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL https://arxiv.org/abs/2010.11929.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018. URL https://arxiv.org/abs/1809.10486.

John Kalkhof, Camila González, and Anirban Mukhopadhyay. Med-nca: Robust and lightweight segmentation with neural cellular automata, 2023. URL https://arxiv.org/abs/2302.03473.

Ailiang Lin, Bingzhi Chen, Jiayu Xu, Zheng Zhang, and Guangming Lu. Ds-transunet:dual swin transformer u-net for medical image segmentation, 2021. URL https://arxiv.org/abs/2106.06716.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL https://arxiv.org/abs/1505.04597.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. URL https://arxiv.org/abs/1409.1556.

Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016. URL https://arxiv.org/abs/1602.07261.

Jeya Maria Jose Valanarasu and Vishal M. Patel. Unext: Mlp-based rapid medical image segmentation network, 2022. URL https://arxiv.org/abs/2203.04967.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. URL https://arxiv.org/abs/1807.10165.