Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)
DOI: 10.54254/2753-8818/5/20230386

Function tissue unit segmentation based on UNext model

Hongyu Sun[1], Qi Zhang[2, 5], Qiangdi Zhang[3] and Zifeng Zhou[4]

[1]Artificial Intelligence and Automation, HuaZhong University of Science and Technology, Wuhan, 430000, China
[2] Faculty of Computer Science, Dalhousie University, Halifax, NS, B3H 4R2, Canada
[3] Software Engineering, Beijing University of Technology, Beijing, Beijing, 100124, China
[4] The Department of Mathematics, University of California, Los Angeles, Los Angeles, 90095, United States

[5] qi.zhang@dal.ca

Abstract. Accurate segmentation for Functional Tissue Units (FTUs) is a challenging issue in past decades. In this study, a model using the dataset of tissue section images will be built to evaluate and mark FTUs across five human organs as clearly as possible. We have the Human Protein Atlas (HPA) as training data and the data from Human BioMolecular Atlas Program (HuBMAP) as testing data. To balance accuracy and inference speed, this study applied Unext, an efficient network based on Unet, as the basic model. We also aim to use some tricks to further improve the performance of the model. First, we used several image enhancement methods to diversify the input image. Second, several structures like Feature Pyramid Network (FPN) and the Atrous Spatial Pyramid Pooling are added to improve model performance and convergence speed. As a result, we successfully segment functional tissue units among images of different sizes. Our proposed model scored 0.56 out of 1.00 by the judge of the competition.

Keywords: machine learning, image segmentation, Unext.

1. Introduction

With the development of medical images, the importance of image segmentation in medical research has gradually emerged. With the use of medical image segmentation, the target region is extracted after being divided into a series of distinct regions with distinct significance. Due to the individual difference and the complexity of human tissue structure, medical image segmentation is difficult to some extent, which deserves more attention.

Functional Tissue Units (FTUs) are cell population neighborhoods that perform an organ's main physiologic function. For example, glomeruli or alveoli is a kind of FTU. Unfortunately, in the average kidney, there are over 1 million glomeruli FTUs, which makes manually annotating FTUs a time-consuming process. Even though there are existing cells and FTU segmentation methods, people are still confronted with many problems: 1) Some methods only focus on one type of organ, which means the segmentation cannot work well when it comes to another organ. 2) Some methods can only use data in the same format. People cannot collect data in the same format, so algorithms should be robust across different dataset differences.

DOI: 10.54254/2753-8818/5/20230386

In the early stages of the deep learning-based image segmentation field, Convolutional Neural Network (CNN), Fully Convolutional Neural Network (FCN) and their modified models were used [1]. These models performed pretty well in medical image segmentation, drawing more attention to this application. Then Ronnebergeret al. brought up U-Net, built on FCN that further improved the performance [1]. Since the U-Net architecture was first proposed in 2015 [2], it has greatly contributed to biomedical image segmentation tasks. Not only can it perform cell segmentation tasks well [2], but also tumor detection [3], red blood cell segmentation with a modified architecture [4] etc. However, U-Net has limitations in learning high-level context and long-range spatial relations due to the inherent locality of convolution operations [5]. Therefore, we introduce Feature Pyramid Networks (FPN) and Atrous Spatial Pyramid Pooling (ASPP) module to our model [6, 7]. FPN is an architecture that takes in an picture to output feature maps at multiple scales, which can be used to detect objects at different levels. ASPP is another way of handling objects at different scales, which uses filters at various sampling rates and effective fields-of-views to probe an incoming convolutional feature layer . In addition, we use U-NeXt, which is a newly designed U-Net architechture, to further improve the performance [8].

In the current stage of medical image segmentation research, U-Net has been used in areas like segmenting cells [2], and a combination of U-Net and FPN has been used in Spine Segmentation [9]. Although there has been work done on improving U-Net in medical image segmentation field, such improvement has not been used in cell-level image segmentation and in fact, U-Net can be further improved.

To begin with, suitable datasets that meet our needs should be prepared. We will use the Human Protein Atlas (HPA) and Human BioMolecular Atlas Program (HuBMAP). Second, we are confronted with is that the data was prepared using a different protocol. That will be challenging for us to adjust our model to function properly. Eventually, when it comes to training, we add several tricks like K-fold Cross Validation, Test Time Augmentation and Model Ensemble to make the model perform well. In this paper, we will differentiate and segment Functional Tissue Units (FTUs) among different kinds of human organs. We are going to set up our model by using a tissue section image dataset, this model must segment FTUs in organs and tissues as accurately as possible. If the result is satisfactory, Relevant researchers' understanding of the relationship between cells and tissues will be improved with our help. When researchers have a deeper understanding of the relationship between cells, they will naturally have a better understanding of the cell functions that affect human health and may make people live a better life.

2. Methodology

2.1. Dataset description and preprocessing

The dataset used in this program is sponsored by two different consortia, the Human Protein Atlas (HPA) [10], and Human BioMolecular Atlas Program (HuBMAP) [11]. The dataset we use consists of data from public HPA data, the test set is a combination of private HPA data and HuBMAP data, and the private test set contains only HuBMAP data. Expect roughly 550 pictures in the hidden test suite, 350 images in the train suite. All HPA images are 3000 x 3000 pixels with a tissue area within the image around 2500 x 2500 pixels. The HuBMAP images range in size from 4500x4500 down to 160x160 pixels. All HPA images have a pixel size of 0.4 μm. For HuBMAP imagery the pixel size is 0.5 μm for kidney, 0.2290 μm for large intestine, 0.7562 μm for lung, 0.4945 μm for spleen, and 6.263 μm for prostate. The sample of train data is demonstrated in Figure 1. The sample of test data is illustrated in Figure 2.

DOI: 10.54254/2753-8818/5/20230386

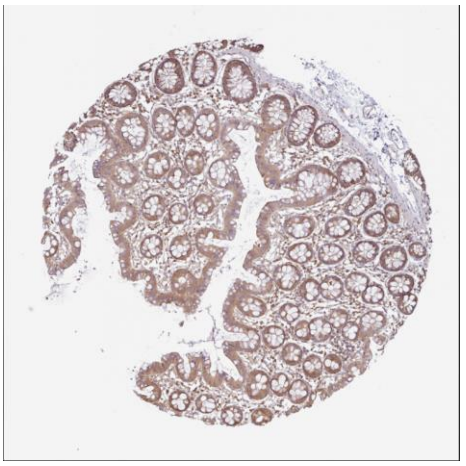

Figure 1. Sample of train data.

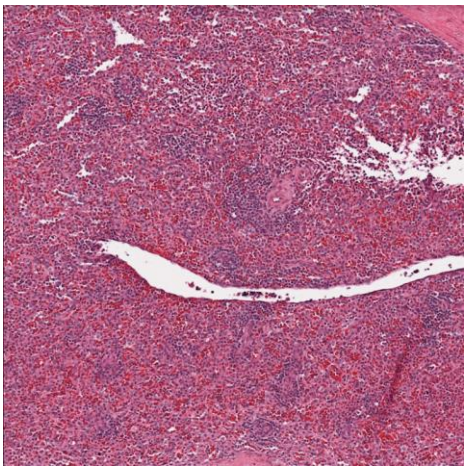

Figure 2. Sample of test data.

Images with different shapes are often considered as one of the major obstacles in image segmentation pre-processing. As mentioned above, the input data is of different size. Therefore, the size of all images is converted into a fixed size of 256×256 for satisfying the requirement of the model's input.

Effective image enhancement always serves as an effective method to improve the models' performance and to improve the convergence speed. In this case, a variety of commonly used image enhancement methods are applied in this study to diversify the image input, including rotating, flipping, scaling, shifting, random enhancement of the image, grid distortion etc.

2.2. Model

The model used in this paper is Unext, a model based on a U-shape network. The model has two parts: namely encoder and decoder. The encoder part creates a representation of features at different levels, while the decoder combines the features and generates a prediction as a segmentation mask. Between the encoder and the decoder, there is a skip connection that allows us to efficiently utilize the intermediate convolutional layer's features from the encoder without running this information completely throughout the encoder and the decoder.

Several key modifications were made to improve the performance. First, it is the Feature Pyramid Network (FPN): between the various upgrade blocks of the decoder and the output layer, an extra skip connection is added. Therefore, the final prediction is generated based on the connection of the Unet output to the intermediate layer output. These jump connections allow gradient flow through a shortcut to improve model capability and convergence rate. Since the middle layer has many channels, their upgrades and as input to the last layer bring significant overhead in computational time and memory. Therefore, a 3x3 + 3x3 convolution is applied before resizing to lower the number of channels.

Second is the Atrous Spatial Pyramid Pooling (ASPP) added between encoder and decoder. The defects of the traditional u-type networks are due to the small receptive fields. Hence, if a model needs to decide on the segmentation of a large object, it can get confused being able to look only into parts of the object. One way to increase the field-of-view and interact between different parts of the picture is to use a block combining different diluted convolutions. While the original paper uses 6,12,18 rates, they may be customized for a particular task and a particular image resolution to maximize the performance. Finally, the decoder upgrade block is based on the pixel shuffling, and not on the transposed convolution used in the first Unet model. It allows for the avoidance of artifacts in the resulting mask.

2.3. Implementation details

Our study is implemented based on Pytorch. We also use fastai, a deep learning library that is

DOI: 10.54254/2753-8818/5/20230386

commonly used nowadays. The implementation process is as follows. First, process the data and create the dataset and dataloader. The input images are preprocessed in the manner described above, and wrapped into the self-defined class HuBMAPDataset. Then we define and build the model. Unext50 with ASPP and FPN is implemented in this section. One of the most commonly used loss function is Lovász Loss, a surrogate of IoU. It can be used when there are only two types of segmentation, that is, only the foreground and background segmentation. In this case, we use Symmetric Lovász loss, a modified Lovász Loss as the loss function. This is to consider not only a predicted segmentation and a provided mask but also the inverse prediction and the inverse mask (predict mask for negative case). Dice Score is used to as an evaluation of our model's performance. We also use K-Fold technique, dividing training dataset into five folds, and each fold is trained for 15 epochs. During the training process, the learning rate gradually increased to a maximum of 0.005, and then gradually decreased with the loss reduction.

3. Result and discussion

3.1. The performance of the model

Table 1. The result from the part of training process.

epoch	train_loss	valid_loss	dice_soft	dice_th	time
1	1.117815	1.220045	0.539429	0.695081	02:11
2	1.091379	1.164033	0.594605	0.719836	02:11
3	1.077110	1.134938	0.637048	0.735647	02:14
4	1.047067	1.212747	0.535532	0.690558	02:08
5	1.028298	1.093837	0.676184	0.768580	02:00
6	0.979540	1.022604	0.669150	0.793170	02:02
7	0.913748	1.034538	0.683080	0.789133	01:53
8	0.879965	1.031025	0.703295	0.800520	01:57
9	0.871885	0.970415	0.701396	0.804225	01:59

According to Table 1, it shows the training process of our model in one fold. The decreasing loss scores and increasing dice scores clearly demonstrate how our model improves gradually.

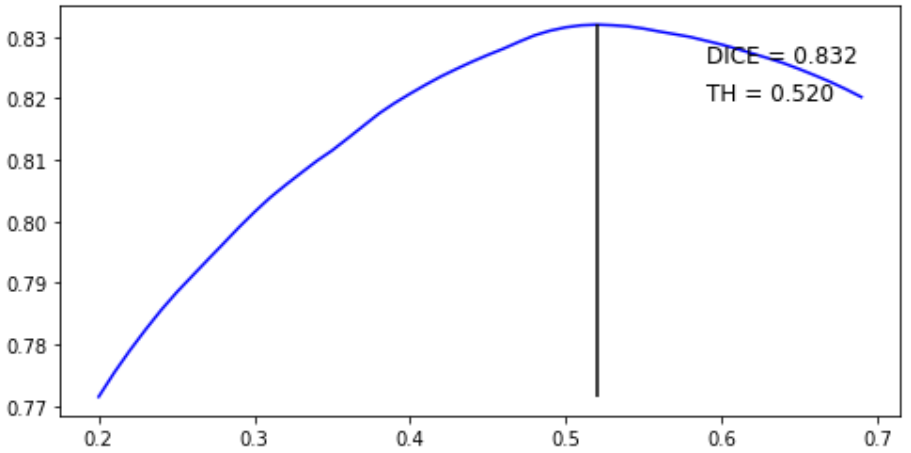

Figure 3. Result of training process.

DOI: 10.54254/2753-8818/5/20230386

Figure 3 shows the relationship between dice score and threshold. After 4 folds, our model achieves a dice score value of 0.832. This figure is using the training data set. It annotates the best dice score achieved by the model and the corresponding threshold value. Eventually, our model is scored 0.56 out of 1.00 by the judge of the competition using a hidden test set. In addition, Figure 4 shows the result of our model's cell segmentation. It identifies and marks cells in the given image.

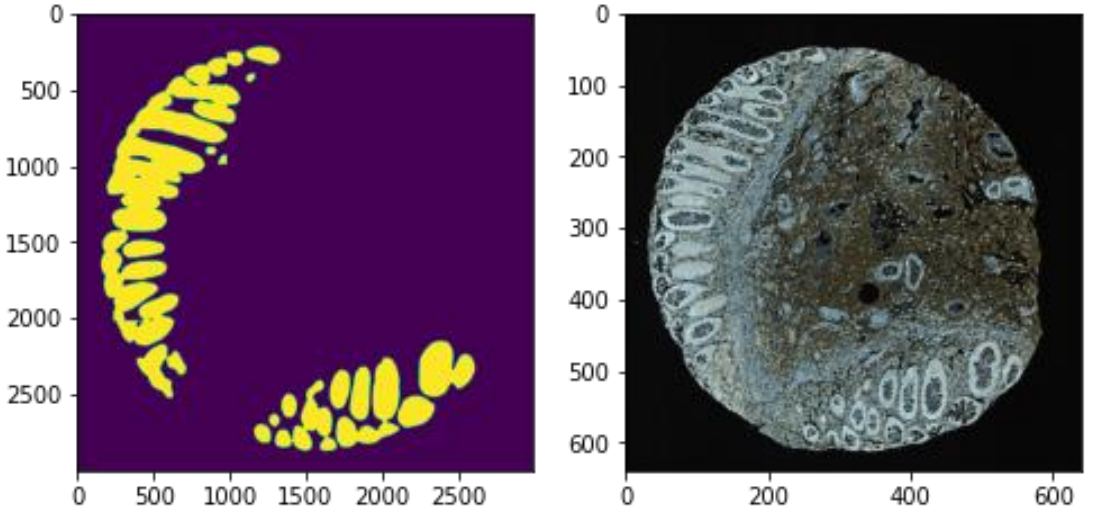

Figure 4. Sample of model segmentation result.

3.2. Discussion

0.56 is a decent score, but it can be better, and we have some ideas that may improve the score. First of all, the image augmentation technique can be used. It does some small changes to the image to become some new images and enlarges the dataset so that the model gets more data to train and has more precise results. Also, using a better pre-trained model improves the accuracy too. We can use the model on some other datasets to train the model first, then we use the model on the dataset so that the model will already be familiar with this type of dataset and make better predictions. By analysis, the data preprocessing part may not be effective enough, there are some more effective data preprocessing can be applied in this model. For example, we can convert the image to multiple smaller images which are similar to image augmentation. Lastly, there is another way to improve the result and it is also our future study topic which is using model swin transformer v2, the most advanced model in medical semantic segmentation. Based on transformer, the model is large enough to accommodate massive data and achieve more decent performance. However, it also requires high computer power such as GPUs to train effectively.

4. Conclusion

In this paper, we build a model to investigate the FTU in different genres of human organs to help improve the world's understanding of relationships between cell and tissue organization. Based on Pytorch, we use the dataset from Human Protein Atlas (HPA) and Human BioMolecular Atlas Program (HuBMAP) and use a portable model Unext to finish our work. Several techniques are also used to improve model performance. Our model is shown to have a dice score of 0.832 on the validation dataset and reach a score of 0.56 by the judge of the competition. The accuracy score is a fair one, but it entails some further work to improve it. One of the reasons for the unsatisfactory score is that we have limited computer power (like GPUs), so we find it difficult to increase the number of training epochs and accommodate a large number of parameters. We come up with several potential solutions such as using the Image Augmentation technique, using a better pre-trained model, and a more

DOI: 10.54254/2753-8818/5/20230386

efficient data preprocessing process. In the future, we will build our model by using swin transformer v2 to improve our model's test results.

Reference

[1] Hesamian M et al. 2019 Deep learning techniques for medical image segmentation: achievements and challenges Journal of digital imaging 32(4) 582-596
[2] Ronneberger O et al. 2015 October U-net: Convolutional networks for biomedical image segmentation In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241) Springer Cham
[3] Amin J et al. 2018 Big data analysis for brain tumor detection: Deep convolutional neural networks. Future Generation Computer Systems, 87, 290-297.
[4] Zhang M et al. 2018 RBC semantic segmentation for sickle cell disease based on deformable U-Net. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 695-702) Springer Cham
[5] Xu G et al. 2021 Levit-unet: Make faster encoders with transformer for medical image segmentation arXiv preprint arXiv:2107.08623
[6] Lin T Y 2017 Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).
[7] Chen L C 2017 Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution and fully connected crfs IEEE transactions on pattern analysis and machine intelligence 40(4) 834-848
[8] Song T 2019 U-next: A novel convolution neural network with an aggregation u-net architecture for gallstone segmentation in ct images IEEE Access 7 166823-166832
[9] Yun D 2022 ELU-Net: An Efficient and Lightweight U-Net for Medical Image Segmentation IEEE Access vol.10 pp.35932-35941
[10] Pontén F et al. 2008 The Human Protein Atlas—a tool for pathology The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland 216(4) 387-393
[11] Lin S et al. 2019 The human body at cellular resolution: the NIH Human Biomolecular Atlas Program.