SEGMENTATION OF CARTILAGE TISSUE IN MICRO CT IMAGES OF MOUSE EMBRYOS WITH MODIFIED U-NET CONVOLUTIONAL NEURAL NETWORK

Jan Matula

Master Degree Programme (2), FEEC BUT

E-mail: xmatul25@stud.feec.vutbr.cz

Supervised by: Jiří Chmelík

E-mail: chmelikj@feec.vutbr.cz

Abstract: Manual segmentation of cartilage tissue in micro CT images of mouse embryos is a very time-consuming process and significantly increases the time required for the research of mammal facial structure development. It is possible to solve this problem by using a fully-automatic segmentation algorithm. In this paper, a fully-automatic segmentation method is proposed using a convolutional neural network trained on manually segmented data. The architecture of the proposed convolutional network is based on the U-Net architecture with its encoding part substituted for the encoding part of the VGG16 classification convolutional neural network pre-trained on the ImageNet database of labelled images. The proposed network achieves average Dice coefficient 0.88 in comparison to manually segmented images.

Keywords: segmentation, cartilage, convolutional neural networks, deep learning

1 INTRODUCTION

X-Ray micro-computed tomography (μCT) is an imaging method capable of capturing image data with high spatial resolution in the range of micrometres. In the research of mammal craniofacial development, mouse animal models are prevalent. The nasal capsule of mouse embryos in their 15th day of development can be measured in the range of millimetres. For this reason, μCT might be considered suitable for non-destructive analysis of these embryos [1]. However, using μCT for analysis of soft tissues also brings some challenges.

An essential step before any further analysis is segmentation of the cartilage tissue. Insufficient contrast between cartilage and surrounding tissues in μCT images makes the use of automatic segmentation algorithms very difficult. In order to achieve the desired accuracy, the cartilage is in most cases being segmented manually by experts. This manual segmentation is a very time-consuming process. In this paper, a fully-automatic segmentation method is proposed using a convolutional neural network trained on a database of manually segmented data. This network segments the cartilage in 2D slices of the 3D CT data of mouse embryo. The accuracy of the proposed automatic segmentation approach is then evaluated in comparison to manually segmented images.

2 CONVOLUTIONAL NEURAL NETWORKS IN IMAGE SEGMENTATION

A commonly used convolutional neural network architecture for binary segmentation of 2D images is the U-Net architecture [2]. It consists of an encoding part and a decoding part. In the encoding part of the network architecture, feature maps extracted from input images are being repeatedly downsampled by using max-pooling layers with ReLU-activated convolutional layers in-between. In the decoding part, the feature maps are upsampled to the size of input by transpose convolution layers. The feature maps from the encoding part are also copied and merged with respective feature maps from the decoding part for better preservation of spatial features of the input images.

3 METHOD

While the U-Net by itself can be trained on smaller datasets, more accurate segmentation can be achieved by using weights pre-trained on large image databases [3]. The encoding part of the U-Net architecture is very similar to the encoding part of VGG16 classification neural network [4]. We used this similarity and substituted the U-Net encoder with VGG16 encoder with its weights pre-trained on the ImageNet database of labelled images [5]. The U-Net decoder was then accordingly modified to be symmetrical to the VGG16 encoder.

A problem that had to be solved in order to train the neural network successfully was the large size of images. The height and width of the segmented μCT slices are in the range of thousands of pixels. To utilise GPU acceleration of the training process, all feature maps extracted from one batch of images must fit in the GPU memory. That makes training on full-resolution images with reasonable batch size impossible on most available hardware.

As a solution to this problem, a downsampling pre-processing step was introduced. First, the voxel intensity values of each sample are standardised with the mean in 0 and standard deviation 1. The neural network architecture requires all input images to be the same size. Because of the variable size of images from different measurements, the images and their respective ground-truth segmentation masks were cropped or padded with zeros to size 1792x1280 pixels. The images were then downsampled four times while the ground truth segmentation masks remained the original size. To compensate for this downsampling step, two additional convolutional blocks were added at the end of the neural network architecture with transposed convolution layers in-between. This means that the final feature maps are upsampled four times and the output of the neural network has the same dimensions as the input before downsampling. This makes training much faster, memory requirements lower and acts as a form of regularisation. The network architecture is shown in **Figure 1**.

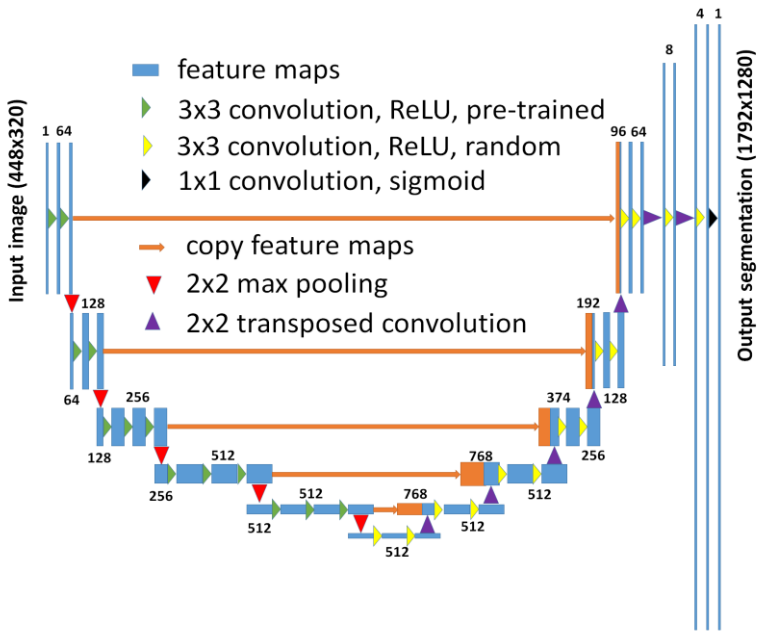

Figure 1: Architecture of the proposed segmentation neural network

Another important matter to consider were augmentations of the training dataset. The micro-CT images of mouse embryos vary in some important aspects. Probably the most substantial is the orientation of the scanned mouse embryo in space. To improve the sample orientation variability in the training dataset, three augmentation techniques were utilised: rotation with a random degree from range from -25° to 25°, horizontal flip and combination of horizontal flip and random rotation. The probability of application of each of the transformations on a particular image is 0.14. With the probability of 0.58, no transformation is applied.

The proposed neural network and all secondary functions were implemented in Python 3.6.5 with Tensorflow-GPU 1.8.0 and Keras 2.2.0 libraries. The training and testing were done on a PC with Intel Core i7 950 CPU, GeForce GTX 980 Ti GPU and 16 GB RAM. The total number of CT slices in the training database was approximately 7000 (around 1200 from each sample). The network was trained with mini-batches of 4 images for 10 epochs. We used Tversky loss function [6] designed specifically for datasets, where the volume of the segmented object is much lower than the volume of background. Adam [7] was chosen as the optimisation algorithm.

4 RESULTS

To evaluate the accuracy of the proposed segmentation method we used a leave-one-out cross-validation method on 7 samples. Dice coefficient was then used for quantitative comparison of the automatically segmented mask and respective ground-truth manual segmentation. The average Dice of all samples is 0.8797 ± 0.0385. **Figure 2** shows contours of both the ground-truth manual and the proposed automatic segmentation in a selected slice. We can see that our neural network follows the cartilage tissue in the CT slice accurately.

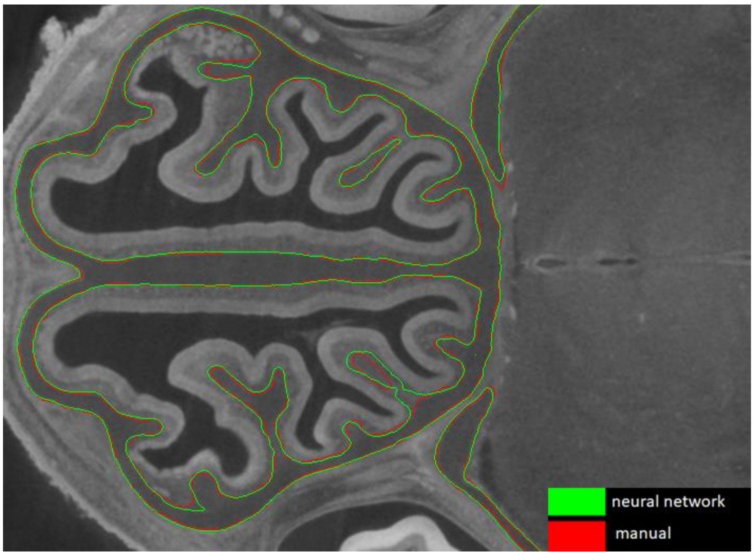

Figure 2: Automatic (green) and manual (red) segmentation of cartilage in a selected slice

5 CONCLUSION

The convolutional neural network architecture developed in this work provides segmentation masks with the accuracy needed for further analysis of cartilage in micro CT images of mouse embryos. This was demonstrated by comparing the results of the automatic segmentation with manual segmentation, where the accuracy achieved is almost 88 % according to the Dice coefficient.

While the segmentation can be considered good, as demonstrated by the high Dice coefficient, there are still some problems to be solved. The segmentation of parts of the cartilage, where the contrast between the cartilage and the surrounding tissue is almost non-existent, is not perfect. Another issue is the number of false positives, which are manifested as specks in the finished segmentation. Both problems might be solved by training the neural network on a larger database of CT images and utilising some more advanced augmentation techniques.

The ability to quickly and automatically segment cartilage tissue in new data will prove useful for fast evaluation of the cartilage development without the need of the time-expensive user input. **Figure 3** shows a 3D model created from 2D binary automatically segmented masks without any further post-processing. The whole segmentation process took roughly 2 minutes on the available hardware, which is a great improvement over the 10 hours it takes an expert to segment the nasal capsule cartilage manually.

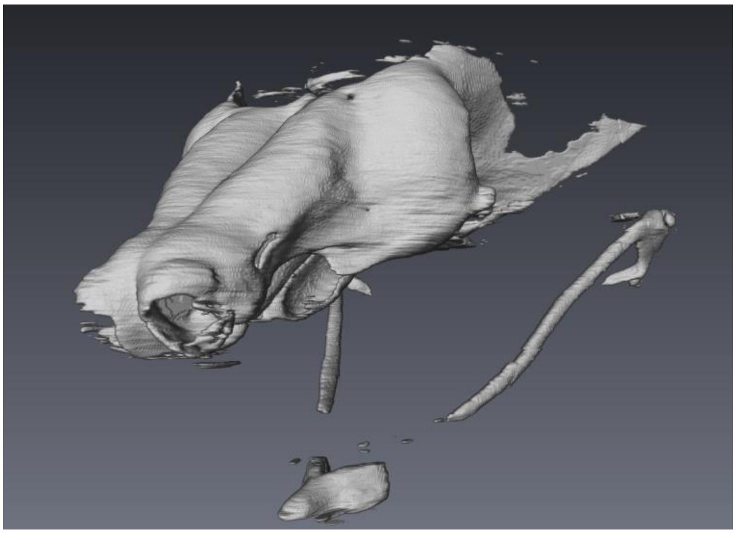

Figure 3: 3D model created from 2D automatically segmented cartilage masks

ACKNOWLEDGEMENT

This research was carried out under the project CEITEC 2020 (LQ1601) with financial support from the Ministry of Education, Youth and Sports of the Czech Republic under the National Sustainability Programme II and CEITEC Nano Research Infrastructure (MEYS CR, 2016–2019).

REFERENCES

[1] Tesařová, M.; Zikmund, T.; Kaucká, M.; et al.: Use of micro computed-tomography and 3D printing for reverse engineering of mouse embryo nasal capsule. *Journal of Instrumentation*, vol. 11, no. 03, 2016: C03006.

[2] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[3] Iglovikov, V.; Shvets, A.: TernausNet: U-Net with VGG11 Encoder Pre-Trained on Image-Net for Image Segmentation. *arXiv preprint arXiv:1801.05746*, 2018.

[4] Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[5] Deng, J.; Dong, W.; Socher, R.; et al.: ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, ISSN 1063-6919, pp. 248–255, doi:10.1109/CVPR.2009.5206848.

[6] Salehi, S. S. M.; Erdogmus, D.; Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging*, Springer, 2017, pp. 379–387.

[7] Kingma, D. P.; Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.