
It doesn't take a whole U-Net to find brain tumors: towards fast brain metastasis segmentation with sparse predictions

Egor Krivov

Skolkovo Institute of Science and Technology
egor.krivov@skolkovotech.ru

Mikhail Belyaev

Skolkovo Institute of Science and Technology
m.belyaev@skoltech.ru

1 Introduction

Medical image analysis has significantly advanced after the popularization of deep learning. For instance, U-Net architecture [1] is currently the state-of-the-art solution for medical image segmentation. It was originally developed for 2D cell image segmentation. However, it is currently applied to 3D image segmentation of brain as well.

In this work, we focus on brain metastasis segmentation from 3D MRI scans. Brain metastases are secondary tumors, which are often treated with stereotactic radiosurgery. This medical procedure currently requires manual delineation of tumors and can apparently be simplified with automatic segmentation. Usually, patients have multiple small metastases, that can be detected on MRI with contrast.

Manual delineation consists of two steps: first detecting metastases by looking through slices, and then attentively delineating particular metastasis. Therefore, doctors do not spend much time analyzing the whole image, focusing attention only on regions with lesions.

However, current state-of-the-art fully-convolutional neural networks (including 3D U-Net architecture) process an image as a whole, without any additional attention to suspicious regions. Therefore, the number of computational operations increase as a third order polynomial from the image size. Deep learning research suggests [2] that deeper networks provide better results, therefore such inefficiency can limit maximum performance.

We argue that it might be effective to split metastasis segmentation into detection and delineation. Detection algorithm would find all regions with lesions so that the second algorithm could perform computationally expensive operations only on suspicious regions. Recently published papers on metastasis segmentation [3, 4] measures performance for both detection and segmentation, suggesting that it is a natural way.

In this paper, we propose a simple network architecture for metastasis detection by predicting downsampled segmentation and compare detection performance between 3D U-Net and our approach. We want to find out if we would miss some regions because of the simpler network architecture and effectively sparse predictions.

2 Method

2.1 3D U-Net

3D U-net that we use as a baseline follows structure from [1]. We replace 2D convolutions with 3D counterparts, following [5]. We also replace convolutions with residual blocks from [6]. We use zero padding to preserve image size and train on big patches from the image ($144 \times 160 \times 128$) with batch size 2. The model was trained for 10 days on a single Tesla M40.

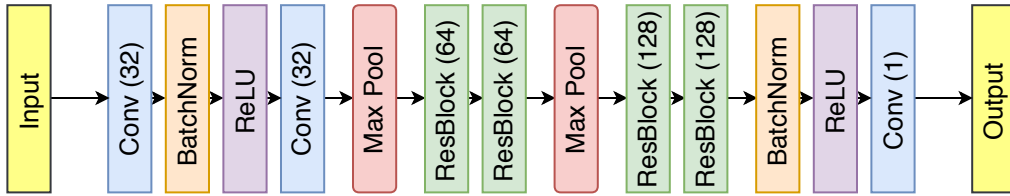


Figure 1: D-Net. All convolutions and performed in 3d. Max pooling is performed with pool size 2

2.2 D-Net

We suggest a simple approach to find metastasis, shown in Fig. 1. Essentially, we just take the beginning of the 3D U-Net encoder, described above. We stop before the third max pooling, ending up with four times downsampled prediction. 3D U-Net architecture would decrease resolution even further and then would restore it with skip-connections, giving the prediction in the original resolution. However, we argue, that it is unnecessary since we are only interested in finding tumors and not dense segmentation.

The purpose of this network is to find all regions with tumors, so we downsample ground truth segmentation with 3D max pooling operation to the shape of the final prediction and use it to compute the binary cross entropy loss. Therefore, the correct prediction for each box 4^3 would be 1 if at least one voxel inside was tumorous.

We call this architecture D-Net since it predicts downsampled segmentation. D-Net doesn't use zero padding. We trained it on patches of size 68^3 as input and 16^3 as output in the original resolution (or 4^3 in the downsampled one), the batch size was 64. Patch size is significantly smaller for D-Net, because it has receptive field smaller than U-Net and, therefore, high patch size is not required. In practice we have noticed, that smaller patch size with higher batch size lead to significantly faster training, probably because of larger tissue diversity in a single batch. We sampled patches in the batch so that a half of them had lesions and the other half did not. Training takes about 22 hours on the same hardware as was used to train U-Net.

3 Experiments

For evaluation, we used a brain metastasis dataset from a Gamma-Knife center. We predict segmentation based on MRI T1c scans. The dataset consists of 404 images, and each image contains four metastasis in average. We used 20% of it as a testing set.

Table 1: Detection results. Sensitivity is a fraction of detected metastases. False positive (FP) is a number of falsely predicted tumors per patient.

	Sensitivity	FP
U-Net	0.771	0.62
D-Net	0.872	1.11

Results are presented in table 1. We consider that metastasis was detected if at least one voxel inside was labeled correctly. An example MRI is presented in Fig. 2

Surprisingly, D-Net works better than U-Net in terms of sensitivity, which is more important for tumor segmentation than the number of false positives. The most probable reason for that is different patch sampling procedure, smaller patch size and larger batch size, used for D-Net, giving it more difficult and diverse examples. However, large patch size, required for U-Net is its intrinsic property, caused by deep encoder structure.

We have also measured average inference time for both models: 1 second for D-Net and 4 seconds for U-Net.

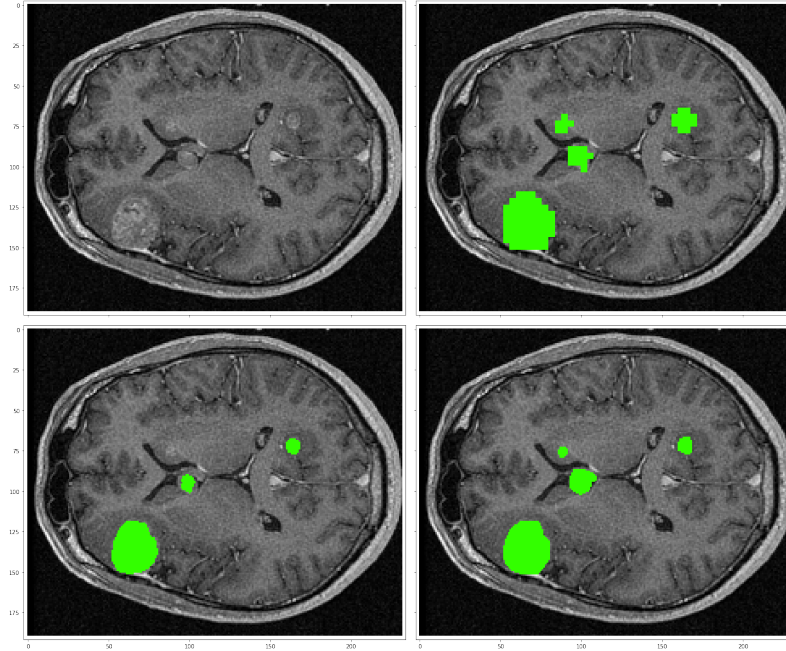


Figure 2: Segmentation example. First row: t1c, D-Net; second row: U-Net, GT

4 Discussion

We have verified that a simple solution can be used for finding metastasis without loss of sensitivity.

Further research would include developing dense segmentation module, that will activate only on suspicious regions, found by the D-Net. Since second part wouldn't have to process the whole image, it could have much more layers than U-Net, potentially improving dense segmentation metrics as well.

References

- [1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [3] Yan Liu, Strahinja Stojadinovic, Brian Hrycushko, Zabi Wardak, Steven Lau, Weiguo Lu, Yulong Yan, Steve B Jiang, Xin Zhen, Robert Timmerman, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS one*, 12(10):e0185844, 2017.
- [4] Odelin Charron, Alex Lallement, Delphine Jarnet, Vincent Noblet, Jean-Baptiste Clavier, and Philippe Meyer. Automatic detection and segmentation of brain metastases on multimodal mr images with a deep convolutional neural network. *Computers in biology and medicine*, 2018.
- [5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016.