Automatic Brain Organ Segmentation with 3D Fully Convolutional Neural Network for Radiation Therapy Treatment Planning

Hongyi Duanmu, SUNY Stony Brook
Jinkoo Kim, Stony Brook University Hospital
Praitayini Kanakaraj, Vanderbilt University
Andrew Wang, Ward Melville High School
John Joshua, Ward Melville High School
Jun Kong, Emory University
Fusheng Wang, Emory University

Journal Title: 2020 IEEE 17TH INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI 2020)
Volume: Volume 2020-April
Publisher: IEEE | 2020-01-01, Pages 758-762
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1109/ISBI45749.2020.9098485
Permanent URL: https://pid.emory.edu/ark:/25593/vphv7

Final published version: http://dx.doi.org/10.1109/ISBI45749.2020.9098485
Accessed November 14, 2022 9:53 AM EST
AUTOMATIC BRAIN ORGAN SEGMENTATION WITH 3D FULLY CONVOLUTIONAL NEURAL NETWORK FOR RADIATION THERAPY TREATMENT PLANNING

Hongyi Duanmu, Jinkoo Kim, Pratayini Kanakaraj, Andrew Wang, John Joshua, Jun Kong, Fusheng Wang

1Department of Computer Science, Stony Brook University
2Radiation Oncology, Stony Brook University Hospital
3School of Engineering, Vanderbilt University
4Department of Biomedical Informatics, Stony Brook University
5Ward Melville High School
6Department of Mathematics and Statistics, Department of Computer Science, Georgia State University
7Department of Biomedical Informatics, Department of Computer Science, Emory University

Abstract

3D organ contouring is an essential step in radiation therapy treatment planning for organ dose estimation as well as for optimizing plans to reduce organs-at-risk doses. Manual contouring is time-consuming and its inter-clinician variability adversely affects the outcomes study. Such organs also vary dramatically on sizes — up to two orders of magnitude difference in volumes. In this paper, we present BrainSegNet, a novel 3D fully convolutional neural network (FCNN) based approach for automatic segmentation of brain organs. BrainSegNet takes a multiple resolution paths approach and uses a weighted loss function to solve the major challenge of the large variability in organ sizes. We evaluated our approach with a dataset of 46 Brain CT image volumes with corresponding expert organ contours as reference. Compared with those of LiviaNet and V-Net, BrainSegNet has a superior performance in segmenting tiny or thin organs, such as chiasm, optic nerves, and cochlea, and outperforms these methods in segmenting large organs as well. BrainSegNet can reduce the manual contouring time of a volume from an hour to less than two minutes, and holds high potential to improve the efficiency of radiation therapy workflow.

Keywords

Deep Learning; Fully Connected Convolutional Neural Network; Brain Segmentation; Radiation Therapy; Biomedical Image Analysis

*The third, fourth and fifth authors performed the work while at Stony Brook University
1. INTRODUCTION

For each external beam radiation therapy treatment, an optimal plan is devised such that the tumor receives the full prescription dose while minimizing the radiation dose to the critical organs nearby. However, to estimate the radiation dose to each organ, the organs need to be delineated on the CT images. Often, the manual delineation is time-consuming (for example, approximately one hour for each head and neck treatment), and the inter-user variability is also one of the fuzzy-factors in clinical outcomes studies. Thus, there is a strong demand for computerized brain organs contouring to support radiation delivery for cancer treatment.

In recent years, with the rapid development of deep learning techniques, many groundbreaking systems emerged for more effective biomedical image analysis tasks [1,2,3] such as urinary bladder [4], liver segmentation [5], brain tumor segmentation [6]. In terms of brain organ segmentation, most of these methods are 2D focused [7], with the limitation on fully taking advantage of 3D contextual information for 3D segmentation. There are mainly three types of methods for 3D brain organ segmentation. 1) “Straight-forward” methods. These are mainly derived from classical image analysis methods such as VGGNet [8], which are built upon stacking several convolutional, pooling or fully connected layers with a specific order. LiviaNet [9] is one representative method. Though those systems are efficient and easy to deploy, they are limited by their model complexity to solve problems with specific challenges. 2) 3D ‘U-Net’. V-Net [10] is one popular 3D version of ‘U-Net’, which provides paired convolutional layers and deconvolutional layers to retrieve the same size of the input image for making predictions of target segmentation. Since it has input image and output image at the same size, it is more suitable for segmenting major objects in large scope and inherently limited on segmentation for images with a mix of large and tiny organs. 3) ‘Multi-module’ methods. Such methods aim at proposing multiple modules with each module focusing on a specific target. For example, FocusNet [11] proposed three different modules – segmentation network, small-organ localization network and small-organ segmentation network to segment target organs in different scales. However, the independence across the modules makes it difficult to optimize the system as a whole.

In this paper, we proposed BrainSegNet, a lightweight, 3D based segmentation framework built upon a fully convolutional neural network, fully taking advantage of 3D contextual information. BrainSegNet takes a weighted loss function and a multi-path architecture combining information at different resolutions for performing segmentation on organs with significant variation in sizes. In particular, it can successfully segment extreme small organs such as cochleae, which can not be segmented with state-of-the-art methods. Instead of constructing a complex system, we demonstrate that this lightweight approach can outperform other advanced systems. In this paper, we focus on segmenting eight organs in brain images: brainstem, eyes (left and right), cochleae (L. and R.), optic nerves (L. and R.), and optic chiasm.
2. MATERIALS AND METHODOLOGY

2.1. Datasets

Among the anonymized datasets of stereotactic radiosurgery treatment plans collected from Stony Brook University Hospital, 46 3D CT volumes were used for this study. All organs were delineated as a part of the clinical process by expert treatment planners. The organs include brainstem, eyes, cochleae, optic nerves, and optic chiasm. The size of the clinical CT volume is $512 \times 512$ for each slice and the number of slices varies from 401 to 569.

2.2. BrainSegNet

BrainSegNet is a 3D fully convolutional neural network (FCNN) consisting of 14 convolutional layers, as shown in Fig 1. To handle the large variations in organ sizes, BrainSegNet includes three paths from the intermediate layers to the final voxel-wise classification for the utilization of information at different scales. To balance the model complexity and the computing burden, we have chosen an input cube size of $32 \times 32 \times 32$. The cube at this size contains enough 3D contextual information and is also suitable to limit the model complexity and computation burden for individual tasks.

**Input and Output.**—As the original raw image is too large to feed into the memory, we crop small $32 \times 32 \times 32$ cubes as input. During the training process, the system will randomly crop cubes, and each cube will contain at least one voxel of the region of interest. During the testing process, it will process small cubes sequentially until the entire image is processed. The output is cubes in size of $8 \times 8 \times 8$ and they will be assembled according to the input cube’s location to form the complete segmentation of the raw image.

**Convolutional layers for image pattern recognition.**—As shown in Fig 1 and Table 1, the first nine layers are the main convolutional layers for image pattern recognition, with a small kernel size of $3 \times 3 \times 3$ chosen for better nonlinear fitting capability [12]. Layers in the same color are grouped as a block as they have the same number of filters. The numbers of filters of these three blocks are 25, 50, and 75 respectively, gradually increased to extract higher levels of image features.

**Fastlane signal layer for multi-layer connection.**—A signal layer ‘fastlane’ is deployed for the multi-path model. It reduces the feature map size of Conv1_3 from $16 \times 16 \times 16$ to $8 \times 8 \times 8$, matching the sizes of the output from layers Conv2_3 and Conv3_3. This will ensure that we can use as much information as possible from different scales and resolutions when contouring, which is essential to small organs contouring.

**Spatially fully connected layers.**—The last four layers are convolutional layers with $1 \times 1 \times 1$ kernels. These convolutional layers are spatially fully connected layers used for making final voxel-wise segmentation decisions, similar to fully connected layers in conventional classification architectures. The number of filters is reduced gradually (400, 200, and 100) preparing for the final voxel-wise classification. The number of filters for the last layer is the same as the number of segmentation targets. In this study, we use nine filters: eight for organs and one for the background.
All convolutional layers are followed by sequentially one batch norm layer [13] and one PReLU layer [14]. For all layers with $3 \times 3 \times 3$ kernel size, one voxel padding is applied to each side of the feature cubes to keep the shape the same between the input and the output. To avoid imbalanced labeling, we designed the weighted cross-entropy loss function as follows:

$$L(\theta) = -\frac{1}{IV} \sum_{i=1}^{I} \sum_{v=1}^{V} \sum_{c=1}^{C} w_c \cdot \delta(y_{vi} = c) \cdot \log \hat{y}_{vc}(x_i)$$

(1)

where $\theta$ denotes all trainable parameters, $y_{vi}$ denotes the ground truth of the $v$-th voxel in the $i$-th image in the dataset, and $\hat{y}_{vc}(x_i)$ is the prediction probability for class $c$ of the $v$-th voxel given the $i$-th raw image. The weighting factor $w_c$ is calculated as follows:

$$w_c = \sqrt{\frac{\max_{c \in C} \sum_{i=1}^{I} N_{ci}^l}{\sum_{c}^{I} N_{ci}^l}}$$

(2)

where $N_{ci}^l$ denotes the total number of voxels of class $c$ in the $i$-th image. This weighting factor normalizes the loss function by the size of class $c$, forcing the system to pay more attention to small targets, to avoid the loss function being overwhelmed by large targets.

2.3. Pre- and Post-processing

**Resizing.**—All raw CT images are resized to $256 \times 256 \times 256$ using bi-linear interpolation to reduce computational and memory requirements.

**Bones Striping.**—The bone voxel values are much larger than those of soft tissue in CT images. Thus, bone striping can make target organs more conspicuous.

**Normalization.**—Normalization has proven to be associated with better segmentation performance. All CT images are normalized by subtracting the mean voxel values and divided by the standard deviation.

**Largest Connected Component (LCC).**—Since the target organs are all single solid objects, the LCC algorithm is used to remove small false positives.

3. EXPERIMENTS AND RESULTS

3.1. Metrics

DSC (Dice Similarity Coefficient) is one of the most used overlap-based metrics in validating medical volume segmentation. DSC increases when the two sets have more overlaps, and it reaches a value of 1 when the two sets are the same.

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2020 August 14.
MHD (Modified Hausdorff Distance) is a modified version of Hausdorff Distance, which is proven to be more robust and gain better ability in discriminatory [15]. Smaller MHD means better similarity, and a zero value MHD means the two sets are the same.

\[
\begin{align*}
MHD &= \max(d(A, B), d(B, A)) \\
d(A, B) &= \frac{1}{N_a} \sum_{a \in A} d(a, B) \\
d(a, B) &= \min_{b \in B} \| a - b \|
end{align*}
\]

where \( A \) and \( B \) denote two different sets (ground truth and segmentation result respectively in our task) and \( \| \cdot \| \) means point-distance measurement (e.g., Manhattan Distance in our studies).

ASD (Average Symmetric Surface Distance) is another commonly used distance metrics, measuring the average of the distance of contouring points between two segmentation results:

\[
\begin{align*}
ASD &= \frac{1}{|A| + |B|} \times \left( \sum_{a \in A} d(a, B) + \sum_{b \in B} d(b, A) \right) \\
d(a, B) &= \min_{b \in B} \| a - b \|
end{align*}
\]

where \( A \) and \( B \) denote two different sets and \( \| \cdot \| \) means point-distance measurement — Manhattan Distance in our studies. Smaller ASD means two sets are closer to each other and while two sets are the same, like other distance metrics, ASD will be 0.

### 3.2. Training Parameters

The dataset consisting of 46 volumes is randomly split into the training set (38 volumes) and the testing set (8 volumes). BrainSegNet was trained over 40 epochs. In each epoch, 10,000 randomly cropped small cubes with corresponding ground truth cubes will be feed into the system for training. The learning rate was initiated with 0.001 and was divided by 10 in the 25th epoch. RMSProp learning strategy [16] was used for training with the loss function of Equation 1. The Nesterov momentum [17] was set to 0.9.

### 3.3. Results

Fig 2 demonstrates example organs segmented. The purple and green objects represent eyes; two optic nerves and optic chiasm are depicted in yellow, blue, and cyan respectively; and two cochleae are depicted in chartreuse green and brown respectively. The biggest object brainstem is shown in red.

We compared BrainSegNet with closely related state-of-the-art methods: V-Net [10] and LiviaNet [9]. They were both fully trained using the same dataset.

Table 2 shows the performance comparisons of different networks. First of all, as shown in the second column, the organ volumes varied significantly. The brainstem was the largest (28.1 cm\(^3\)), and it was approximately 200 times larger than the smallest organ cochlea (0.14 cm\(^3\)).
It was evident that the performance was impacted by the organ size. All methods, except BrainSegNet, failed to segment the cochleae.

Though both V-Net and LiviaNet are networks widely proven to be efficient and achieved similar performance in large organs (brainstem and eyes), with no specific design for small organ segmentation, they perform poorly when organ size becomes smaller. Even for cochleae, both failed to generate any meaningful segmentation results. Our system BrainSegNet by using multi-layer connection and weighted loss function was able to segment organs at all scales. It achieved the best performance in nearly all metrics.

Tab 3 shows a comparison from the perspective of model complexity. V-Net, a 3D version of U-Net, is naturally more complicated than a ‘straight-forward’ system, like LiviaNet or our BrainSegNet. By elegant modification, BrainSegNet can perform better without much burden on computation and storage, making it more practical for real clinical usage. It proves that accurate modification can solve the tasks well without endlessly stacking layers or trying numerous advanced architectures. Also, BrainSegNet can segment one CT volume in two minutes on one NVIDIA Tesla K80. V-Net is chosen as a representative ‘U-Net’ systems, and LiviaNet is chosen as a typical straight-forward system for our comparative studies. A recent ‘multi-module’ based system [11] combines multiple independent models, which creates higher complexity, and is more difficult to optimize. Due to the unavailability of source codes for commercial license reason, we did not include it for comparison.

In this study, the ground truth was from manual segmentation by more than one treatment planners. Therefore, there was inter-observer variation in the reference dataset to some extent. In particular, we expect there were more uncertainties in the optic nerve and chiasm segmentation because the visibility of their boundaries is very low on CT images. Especially, the optic chiasm is not visible directly on the CT images. It needs to be contoured with registered diagnostic magnetic resonance (MR) images. We speculate that it was one of the reasons for the low DSC value (43 percent) of the optic chiasm segmentation (Table 2), in addition to their small organ volumes. Also, Fig 3 shows one example of cochleae contouring results, where red and yellow ones are predictions, and orange and purple ones are delineated by experts. While the metrics for these tiny organs do not show perfect values, the visual evaluation demonstrates that the segmentation can precisely locate these organs.

For radiation therapy treatment planning, the results of this study are very promising. When an organ to be contoured is distant from the treatment target(s), the slight inaccuracy in contouring is tolerable. The results in this study are suggestive that BrainSegNet can potentially reduce or eliminate the manual segmentation for treatment planning and improve the clinical process. One of the ongoing work is a pilot study to integrate BrainSegNet into Varian’s Eclipse, a popular treatment planning system used in cancer centers.

4. CONCLUSION

Automated contouring has the potential to significantly reduce or eliminate the time-consuming manual contouring for radiation therapy treatment planning. BrainSegNet is built on a fully connected convolutional neural network designed to incorporate information at
multiple scales in 3D and avoid imbalanced labeling with an optimized loss function. BrainSegNet outperforms other state-of-the-art methods on segmenting normal organs and can segment tiny organs at the same time, which are not possible with other methods. Future studies will be on exploring a larger clinical datasets and applying them to other organs in the body.

**ACKNOWLEDGEMENT**

This research is supported in part by grants from National Institute of Health 7K25CA181503-06, 1U01CA242936-01, and National Science Foundation ACI 1443054 and IIS 1350885.

**6. REFERENCES**


Fig. 1.
A flow chart of BrainSegNet architecture. Rounded rectangles denotes convolutional layers grouped by color. The circle denotes concatenating feature maps along the channel dimension.
Fig. 2.
Example segmented 3D organs from BrainSegNet.
Fig. 3.
An example of cochleae contouring results. Red and yellow: segmentation results; orange and purple: annotations.
Table 1.
The architecture details of BrainSegNet.

<table>
<thead>
<tr>
<th>Layer Name</th>
<th>Input Shape (# of channels)</th>
<th>Output Shape (# of channels)</th>
<th>Kernel Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1_1</td>
<td>32 × 32 × 32 (1)</td>
<td>32 × 32 × 32 (25)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv1_2</td>
<td>32 × 32 × 32 (25)</td>
<td>32 × 32 × 32 (25)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv1_3</td>
<td>32 × 32 × 32 (25)</td>
<td>16 × 16 × 16 (25)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv_Fastlane</td>
<td>16 × 16 × 16 (25)</td>
<td>8 × 8 × 8 (25)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv2_1</td>
<td>16 × 16 × 16 (25)</td>
<td>16 × 16 × 16 (50)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv2_2</td>
<td>16 × 16 × 16 (50)</td>
<td>16 × 16 × 16 (50)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv2_3</td>
<td>16 × 16 × 16 (50)</td>
<td>8 × 8 × 8 (50)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv3_1</td>
<td>8 × 8 × 8 (50)</td>
<td>8 × 8 × 8 (75)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv3_2</td>
<td>8 × 8 × 8 (75)</td>
<td>8 × 8 × 8 (75)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv3_3</td>
<td>8 × 8 × 8 (75)</td>
<td>8 × 8 × 8 (75)</td>
<td>3 × 3 × 3</td>
</tr>
<tr>
<td>Conv4_1</td>
<td>8 × 8 × 8 (150)</td>
<td>8 × 8 × 8 (400)</td>
<td>1 × 1 × 1</td>
</tr>
<tr>
<td>Conv4_2</td>
<td>8 × 8 × 8 (400)</td>
<td>8 × 8 × 8 (200)</td>
<td>1 × 1 × 1</td>
</tr>
<tr>
<td>Conv4_3</td>
<td>8 × 8 × 8 (200)</td>
<td>8 × 8 × 8 (100)</td>
<td>1 × 1 × 1</td>
</tr>
<tr>
<td>Conv4_4</td>
<td>8 × 8 × 8 (100)</td>
<td>8 × 8 × 8 (9)</td>
<td>1 × 1 × 1</td>
</tr>
</tbody>
</table>
## Table 2. Performance comparison

<table>
<thead>
<tr>
<th>Organs</th>
<th>Volume (cm$^3$)</th>
<th>DSC</th>
<th>MHD (mm)</th>
<th>ASD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brainstem</td>
<td>28.1</td>
<td>0.77</td>
<td>0.32</td>
<td>0.59</td>
</tr>
<tr>
<td>L. Eye</td>
<td>8.15</td>
<td>0.82</td>
<td>0.78</td>
<td>0.31</td>
</tr>
<tr>
<td>R. Eye</td>
<td>7.96</td>
<td>0.84</td>
<td>0.76</td>
<td>0.16</td>
</tr>
<tr>
<td>Optic Chiasm</td>
<td>0.84</td>
<td>0.32</td>
<td>0.36</td>
<td>0.15</td>
</tr>
<tr>
<td>L. Optic Nerve</td>
<td>0.81</td>
<td>0.53</td>
<td>0.52</td>
<td>0.15</td>
</tr>
<tr>
<td>R. Optic Nerve</td>
<td>0.82</td>
<td>0.56</td>
<td>0.56</td>
<td>0.20</td>
</tr>
<tr>
<td>L. Cochlea</td>
<td>0.14</td>
<td>0.56</td>
<td>0.70</td>
<td>0.30</td>
</tr>
<tr>
<td>R. Cochlea</td>
<td>0.14</td>
<td>0.56</td>
<td>0.64</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Methods compared include V-Net (VN), LiviaNet (LN) and the proposed BrainSegNet (BSN). The measurements are averaged on all volumes in the testing set. ‘NaN’ means no segmentation for this class is generated.
Table 3.

Model complexity comparison

<table>
<thead>
<tr>
<th>Model</th>
<th># parameter</th>
<th># floating point operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-Net</td>
<td>$4.12 \times 10^7$</td>
<td>$3.1 \times 10^{11}$</td>
</tr>
<tr>
<td>LiviaNet</td>
<td>$7.80 \times 10^5$</td>
<td>$3.6 \times 10^9$</td>
</tr>
<tr>
<td>BrainSegNet</td>
<td>$7.86 \times 10^5$</td>
<td>$2.8 \times 10^9$</td>
</tr>
</tbody>
</table>