# Instance Tumor Segmentation using Multitask Convolutional Neural Network

Mina Rezaei, Haojin Yang, Christoph Meinel
*Hasso Plattner Institute, University of Potsdam*
Email:{mina.rezaei,haojin.yang,christoph.meinel}@hpi.de

***Abstract*—Automatic tumor segmentation is an important and challenging clinical task because tumors have different sizes, shapes, contrasts, and locations. In this paper, we present an automatic instance semantic segmentation method based on deep neural networks (DNNs). The proposed networks are tailored to picture tumors in magnetic resonance imaging (MRI) and computed tomography (CT) images. We present an end-to-end multitask learning architecture comprising three stages, namely detection, segmentation, and classification. This paper introduces a new technique for tumor detection based on high-level extracted features from convolutional neural networks (CNNs) using the Hough transform technique. The detected tumor(s), are segmented with a set of fully connected (FC) layers, and the segmented mask is classified through FCs. The proposed architecture gives promising results on the popular medical image benchmarks. Our framework is generalized in the sense that it can be used in different types of medical images in varied sizes, such as the Liver Tumor Segmentation (LiTS-2017) challenge, and the Brain Tumor Segmentation (BraTS-2016) benchmark.**

## I. Introduction

Medical imaging plays an important role in disease diagnosis, treatment planning, and clinical monitoring. Image segmentation is a fundamental problem in medical image analysis, which attempts to identify the exact boundaries of objects such as organs or abnormal regions (e.g. tumors) based on images. The diversity of medical image acquisition methods (e.g. MRI, CT, ultrasound) and geometry (2D vs. 3D) can result in variations in the appearance of body organs and tumor shape [1], thus making segmentation more challenging. Hence, manual segmentation can be very time-consuming and subjective, and an accurate semantic segmentation algorithm based on different medical images could help to improve prediction accuracy, and efficiency, and assist in treatment planning.

Over the past few years, numerous automatic approaches have been developed to speed up medical image segmentation based on deep neural networks (DNNs). DNNs can learn hierarchical representation of image data without requiring any effort to design handcrafted features [2]–[4]. Recent studies, focus on the standard segmentation problem, which involves assigning class labels to each pixel in an image, but they say nothing about the number of instances of each class [5]–[8]. Instance segmentation poses the problem of identifying individual instances of objects and their categories. Instance segmentation is very important in a number of applications, such as medical imaging, autonomous driving, image captioning, and visual question answering and each application that we need to count and specify same objects.

In this paper, we investigate the applicability of CNNs in tumor segmentation based on different medical image modalities. Our goal is to perform localization, segmentation, and instance classification in volumetric clinical images (e.g. CT and MRI). To this end, we propose a single multitask framework. Contrary to previous multitask learning studies [9], we propose a new integrated architecture where the later stage relies on the output of the earlier stage. The multitask that we address in this paper includes tumor detection, segmentation, and instance classification. Furthermore, we introduce a new technique for bounding box detection, namely Hough-boxes, using Hough transform features, in combination with high-level extracted features from CNNs. All proposed regions are segmented through fully connected neural networks. Instance classification labels are predicted based on shared features from previous stages and fully connected models.

We provide a detailed evaluation of parameter variations and network architectures. The contribution of this work can be summarized as follows:

- We propose a new multitask learning architecture for instance segmentation and obtain promising results on two different medical image benchmarks. The achieved accuracy in terms of Dice is 0.82 for whole brain tumor segmentation on 2016 BraTS and 0.74 for segmentation of liver tumor(s) on LiTS-2017.
- We propose a new region proposal technique using the Hough transform features on top of depth convolutional features. We share our implemented code [1] with communities for further research use.
- We introduce a new mini-batch normalization, namely a patient-wise technique. We experimentally show that it helps to accelerate the learning process and improve the accuracy.

The rest of the paper is structured as follows: Section II gives a brief description of some related approaches, Section III describes the propose method, Section IV presents experimental results in detail, and Section V concludes the paper.

[1] https://github.com/MedInstSeg/

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

## II. Related Work

The number of publications and clinical studies on automated medical image segmentation, e.g. brain tumor segmentation [10] and liver tumor segmentation [6], using deep learning has increased significantly in the last few years. The most common architectures for task of segmentation can be categorized into four general types, namely fully convolutional networks (FCNs), recurrent neural networks (RNNs), hybrid models, and detector-based approaches.

Ronneberger *et al.* [7] presented a fully convolutional neural network with restricted connection between layers, namely U-net, for segmenting neuronal structures in electron microscopic stacks. The Unet has achieved very good performance in medical image segmentation. Later, 3D-Unets were used for brain tumor segmentation [11], and also for liver segmentation [12].Beers *et al.* [13] used sequence of 3D-Unet for brain tumor segmentation and in different architectures with conditional random fields [6] for liver segmentation.

Other studies have presented models composing of different RNN units (one per class) which receive a spatial class score map and learn to separate the different instances that are included in the input [14], [15]. In medical image segmentation, sequential convolutional LSTMs [16] take the sequential frame of 2D images and give separate object segments as outputs.

The convolutional neural network cascade and hybrid models are other successful categories. The InputCascadeCNN, introduced by Havaei *et al.* [17], has obtained good performance for brain tumor segmentation. The Input-CascadeCNN has image patches as inputs and uses a cascade of CNNs in which the output probabilities of a first-stage CNN are taken as additional inputs to a second-stage CNN. Christ *et al.* [6] proposed a 3D CNN using two pathways with inputs of different resolutions. 3D CRFs were also needed to refine their results.

Detector-based approaches [18], [19] first predict bounding box proposal and then perform segmentation within each region of interest (ROI). Similarly [18], our proposed method learns both the detector and the segmentor but using different strategy for selecting bounding boxes. We use Hough transform features in addition to the high-level features extracted from CNNs as an accurate localization technique.

Milletaria *et al.* [20] presented a CNN architecture, in combination with a Hough voting approach inside of convolutional layers for brain image segmentation. However, their method is not end-to-end and only works for compact blob-like structures. The architecture and multitask learning are the other key differences between our proposed method and that of Milletaria *et al.* [20], who only addressed a segmentation task.

## III. Methodology

In this paper, we design a new multitask architecture to solve the problem of instance semantic segmentation. As shown in Fig. 1, the proposed framework is divided into three steps, namely region proposal detection, object segmentation, and instance classification. Each step has a loss, but the three losses are not independent, and a loss of later stage relies on the output of an earlier stage. We provide more details about the region proposal technique in Section III-B, the tumor mask segmentation in Section III-C, and instance classification in Section III-D. The proposed architecture in this paper follow a patient-wise, mini-batch normalized method as described in Section: III-A.

### A. Patient-Wise Batch Normalization:

Of late, several popular techniques have been developed for normalization such as batch normalization [21] and max norm constraints [22] with core idea being shifting the inputs to a zero mean and unit variance. The inputs are normalized before applying the non-linearity in order to prevent the inputs from saturating extreme non-linearity. As shown by Ioff *et al.* [21], batch normalization has improved overall optimization and gradient issues. In many cases, initial weights have a large deviance from true weights. These outliers need to be compensated by the gradients, which increase the delay in convergence during training. The batch norm reduces the influence from weight deviance by normalizing the gradients which make training faster.

We initially normalize the inputs where the mean and variance are computed on a specific patient from the same acquisition plane (Saggital, Cronal,and Axial) and from all available image modalities (e.g. t1, t1ce, t2, and Flair in the BraTS benchmark). With this progress, the deviances get increasingly larger, and our back-propagation will need to account for these large deviances. This restricts us from using a small learning rate to prevent gradient explosion. For example, the mini-batch with 128 images includes the same patient images from four available modalities from the same acquisition plane. Algorithm 1 shows how to compute normalization at each mini-batch by patient-wise batch normalization.

**Algorithm 1:** Patient-wise mini-batch normalization

**Input :** Values of $x$ over a mini-batch: $\beta = x_1, x_2, ...$
Parameters to be learned:$\gamma, \beta$
**Output:** $y_i = BN_{\gamma,\beta}(x_i)$

**for** $Patient : P_1, P_2, ..., P_n$ **do**
  **for** $AcquisitionPlane : x_i, y_i, z_i$ **do**
    **for** $ImageModalities : t1, t2, t1ce, Flair$ **do**
      $\mu_\beta \leftarrow \frac{1}{m}\sum_{i=1}^{n} x_i$
      $\sigma_\beta^2 \leftarrow \frac{1}{m}\sum_{i=1}^{n}(x_i - \mu_\beta)^2$
      $\hat{x_i} \leftarrow \frac{x_i - \mu_x}{\sqrt[2]{\sigma_x^2 + \varepsilon}}$
      $y_i \leftarrow \gamma\hat{x_i} + \beta = BN_{\gamma,\beta}(x_i)$
    **end**
  **end**
**end**

### B. Region Proposal Detection:

3D bio-medical images are often represented as the sequence of 2D slices (such as z-stacks). The proposed archi-

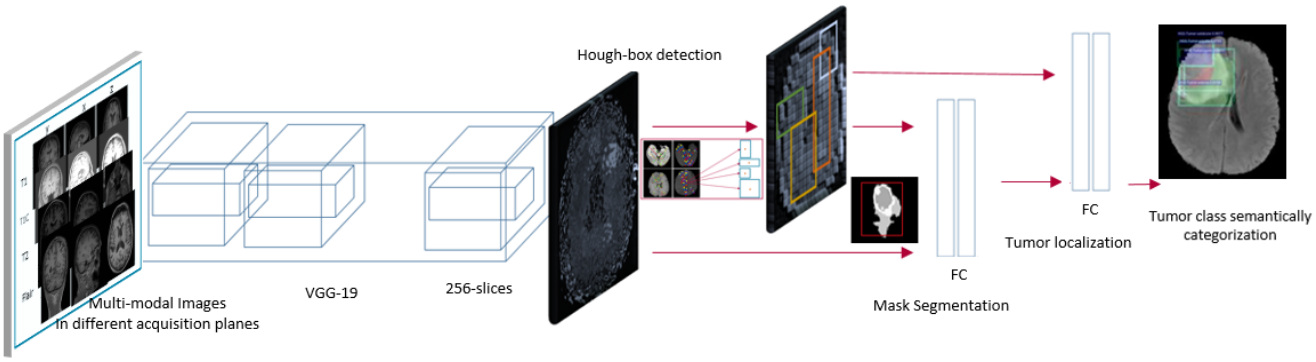


Fig. 1. The multitask CNN diagram for semantic, and instance tumor segmentation. The proposed framework has three stages, named detection, segmentation, and classification. Our framework takes multi-modal images as an input and outputs, semantic mask of tumors. As first stage, bounding box proposal predicted on top of CNN network (e.g. VGG-19) using novel approach of Hough box technique. Predicted boxes get segmented through CNNs in second stage. The tumor masks categorized and counted in the last stage.

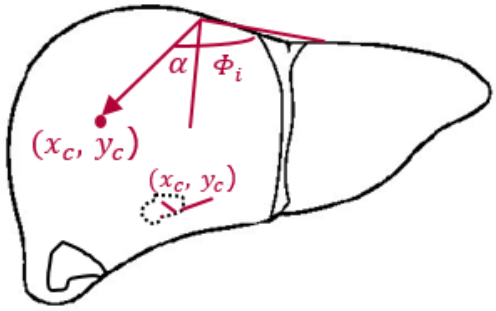


Fig. 2. An example geometry shape of liver from LiTS dataset that we used to form the Table 1.

tecture takes 2D slices with various sizes as inputs. In the first step, the network is trained to predict bounding boxes around the tumor(s). In our experiments, we investigate the VGG-19 style [4] architecture as the feature extractor, which has 13 shareable convolutional layers. The proposal sampling process is performed on top of the dense feature layer after conv5-4. To generate region proposals, we follow the solution suggested by Ren *et al.* [23], which puts the small network over the conv feature map output by the last shared conv layer. This network is fully connected to an $n \times n$ spatial window of the input convolutional feature map. Each sliding window is mapped to a lower dimensional vector (512d). Unlike the faster-RCNN [23] in our architecture, this vector is fed to only the bounding box regression layer. To continue, we choose center of ROIs based on the Hough transform feature while faster-RCNN selected $k$ randomly center. Since most tumors (in liver, brain, lung, etc.) have curved or circular geometric shapes, we are motivated to find a center of curve on the images. We regress bounding boxes around each curve center. In the rest of the paper, we call the Hough anchor as the center of ROIs. The proposed method selects the position invariant of the Hough anchor into the following steps, as described in B.1.

TABLE I
INCREMENTATION IN THE GENERALIZED HOUGH TRANSFORM

| Angle measured from image boundaries to reference point | Set of radius $r^k$ where $r = (r, \alpha)$ |
|---|---|
| $\phi_1$ | $(r_1)^1, (r_2)^1, ..., (r_n)^1$ |
| $\phi_2$ | $(r_1)^2, (r_2)^2, ..., (r_n)^2$ |
| ... | ... |
| $\phi_m$ | $(r_1)^m, (r_2)^m, ..., (r_n)^m$ |

*B.1. Position-Invariant Hough Anchors:* Hough transform is a global method for straight lines and circle detection in classical image processing. Here, we adopt the Hough transform method, as described in [24] to find the centers of curved line in medical images. At each sliding windows, we classify a given image into foreground and background to find an edge point $F(x, y)$ (e.g. by applying the Canny edge detector [25]). Then each edge point $(x, y)$ is mapped to the Hough space to create an accumulator array and to find the local maximum.

We follow the described step in Algorithm 2 to find the center $(x_c, y_c)$ curved shape with radius $r$. As described in Table 1, parameters $\phi$ and $r$ are set.

Then the center points are selected with a specific distance from the detected curve line; For example we choose four different ratios. Each anchor is centered at the sliding window, and is associated with a scale and an aspect ratio. Fig. 3, right column illustrates the different ratio boxes surrounding the Hough anchors. Subsequently, each output feature vector is further fed into a sequence of fully connected layers, which is for mask segmentation.

*C. Tumor Mask Segmentation:*

In our work flow, the second task is pixel-level segmentation. The network is trained on the shared convolutional features from previous layers for the segmentation of each proposed box. To come over on heterogeneous collection of computed bounding boxes, we follow the region of the interest pooling layer [26] as solution and divide them into

**Algorithm 2:** Hough transform technique for finding possible center of curved shapes.

**Input :** The filtered image $F(x, y)$ and a candidate position $(x, y)$

**Output:** Possible center locations $(x_c, y_c)$ for the shape are given by maxima in accumulator array

**Step 1:** Make Table 1, to find the shape locations.

**Step 2:** From an accumulator array of possible reference point $A(x_{cmin} : y_{cmin}, x_{cmax} : y_{cmax})$

**Step 3: for** $EdgePoint : (x_1, y_1), (x_2, y_2), ..., (x_i, y_i)$ **do**

  Step 3a: Compute $0 \leq \phi(i) \leq 180$.

  Step 3b: Calculate the possible centers $(x_c, y_c)$ :

$$x_c = x + r\phi cos(\alpha(\phi))$$

$$y_c = y + r\phi sin(\alpha(\phi))$$

  Step 3c: Increment the accumulator array

$$A(x_c, y_c) = A(x_c, y_c) + 1;$$

**end**

**Step 4:** Possible center locations for the shape are given by maxima in accumulator $A$.

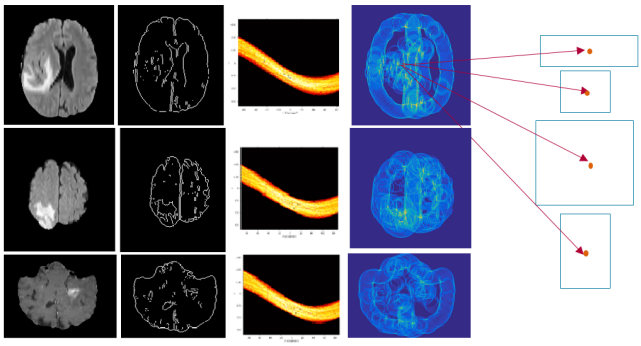

Fig. 3. From left to right: the first column shows the 2D images in an arbitrary size as an input to the architecture, second column shows binary images after canny edge detection. The third column shows the Hough accumulator space, fourth column is curve line and middle point detected by Hough transform. In the last column, high scored center points (Hough anchor) are selected for center of the sliding window surrounding the objects.

a pyramid grid of sub-windows. Here three pyramid levels, namely $4\times4$, $2\times2$, $1\times1$ and average pooling have been applied in each sub-window to generate the fixed-size vector as an input for fully connected layers. We use two fully-connected layers in this step. The first fully-connected layer (with ReLU) reduces the dimension to 256, followed by the second fully-connected layer that regresses a pixel-wise mask. This mask is the network output, representing a list masks from logistic regression output (via sigmoid) with continuous values in $[0, 1]$. As related work, DeepMask [27] and MNC [18] also regress discretized masks. The DeepMask applies the regression layers to dense sliding windows (fully-convolutionally), and MNC [18] regresses masks from $k$ randomly selected boxes. Our method only regresses masks from a few proposed boxes with the Hough anchor center.

### *D. Tumor Mask Classification:*

The final task of the proposed architecture is instance categorization where prediction is on top of two previous steps. The pixel-level masks, together with the shared convolutional features of boxes are the inputs into fully connected layer in the last stage. We concatenated the box vector and the mask vector on top of a softmax classifier with $n+1$ ways to predict $n$ categories plus one background category.

## IV. Experiments and Results

We have tested the proposed architecture on real patient data from two popular medical image segmentation benchmarks. We evaluated the ability of our methods in semantic segmentation as well as instance segmentation.

### *A. Datasets:*

*Brain Tumor Segmentation:* The BraTS2016 benchmark [10], prepared the data in two parts, namely high and low grade glioma (HGG/LGG) brain tumor(s), in magnetic resonance images (MRI). All brains in the dataset had the same orientation. For each brain there were four modalities t1, t1ce, t2, and Flair all of which were co-registered. The training brains came with ground truth for which three segmentation labels were provided necrosis, enhancing tumor, edema. The BraTS2016 released data in three sub sets train, validation and test respectively consisting of 289, 47, and 147 MRI images in four modalities, which the annotated file provided only for the training set.

*Liver Tumor Segmentation:* The LiTS2017 benchmark is a real-life clinical computer tomography (CT) dataset from multiple CT scanners and acquired at different centers. The examined patients were suffering from different kinds of cancerous diseases with different manifestations in the liver. The training dataset contained 130 CT scans while the test dataset had 70 CT scans. The ground truth by medical experts provided all training samples with three segmentation labels, namely liver, lesion(s) of the liver, and background.

We consider 20% of the data for testing and 80% for training time.

### *B. Data Preparation:*

Since we aimed to explore the proposed method on two different medical datasets, we provide different pre-processing algorithms. The grayscale distribution of MR images is dependent on the acquisition protocol and the hardware. This makes learning difficult since we expect to have the same data distribution from one subject to another. Therefore, pre-processing is an important step to bring all subjects in similar distributions.

We applied a bias field correction on the MR images to correct the intensity non-uniformity in the MR images using N4ITK [28]. Lastly, we applied histogram matching

normalization on the MR images. For the CT images, pre-processing is carried out in a slice-wise fashion. We applied Hounsfield unit (HU) values, which were windowed in the range of $[100, 400]$ to exclude irrelevant organs and objects. In next step, we applied histogram equalization to increase the contrast for better differentiation of abnormal liver tissue.

Fig. 4, the first row in second column shows the final slice after Hounsfield unit windowing and contrast enhancement. The second row of Fig. 4 shows an input example ready for training process from BraTS2016.

### C. Implementation and Configuration:

Our framework is implemented based on the modified open-source deep learning library Caffe [29]. As mentioned before, we designed an end-to-end architecture for multitask learning. The proposed architecture takes all available modalities from the same patient, same acquisition plane, and same slice as an input and an output is the instance semantic segmented mask of tumor. We share our implemented code in MATLAB2016b for instance data preparation.

The first stage is detection when the data is trained on the VGG-19 network style. On the top of the conv5-4, we used the same network introduced by faster-RCNN [18] namely region proposal network (RPN). As opposed to the RPN proposed by Ren *et al.* [23], we selected the initial bounding box center using the Hough transform technique for two reasons: first to prevent the generation of 10k random boxes; and second, since most tumors have geometrically circular shapes, we wanted to predict the center tumor through Hough transform curved shape detection. Fig. 5, in the right column shows the accumulator space and curved line in the CT of liver images. The Hough transform produces $\sim$ 70 center points (Hough anchor). Like Ren *et al.* [23], we used four aspect ratios (0.25, 0.5, 1, and 1.5) for generating bounding boxes with m $\times$ n surrounding the Hough anchors. We set window sizes as $m = n = 16$ in the BraTS benchmark and it is 8 in the LiTS benchmark. Then, in the first stage network we have $\sim 70 \times 4$ regressed boxes. We use non-maximum suppression (NMS) to reduce redundant candidates. The threshold of the intersection-over-union (IoU) ratio for this NMS is 0.75 like MNC [18]. The implemented region proposal detection by Hough-box is available on the Github [2].

After that, the top-ranked 40 boxes are used in the segmentation stage. In the second stage, the segmentation mask is computed from the intersection of the Hough boxes and the ground truth mask. Like MNC, we also consider two sets of positive, and negative samples at this stage. In the first set, the positive samples are the instances that overlap with ground truth boxes by Hough box when intersection of union (IOU $\geq$ 0.5) and segmented mask IoU $\geq$ 0.5. The loss function of last stage has $(n + 1)$ classifiers.

All training and experiments were conducted on a workstation equipped with a single NVIDIA TITAN X GPU. The learning rate is initially set to 0.001 and the network is trained with stochastic gradient decent implemented by the Caffe library.

### D. Evaluation and Discussion:

In the training phase for brain tumor segmentation, our network takes four MRI image modalities as inputs and learn three masks of whole-tumor, enhanced-tumor, and core-tumor (four-inputs into three-outputs) as shown in Fig. 6. Table. 4 shows the semantic region of tumors predicted by the proposed method in test time in detection and segmentation. We also compared our results with the recent related approaches as shown in Table 2 and 3, where the reference method also used LITS datasets. For liver lesions segmentation, our network gets one image modality (CT-images) as an input and learn the instance mask of lesions (one-input into one-output). Fig. 7 shows the instance segmentation results for liver lesions. Its worth to mention, we trained Unet and MNC network on the same training and testing data as ours. Regarding the comparison results from Table 2, 3, and 4, we can see the advantage of patient-wise batch normalization and higher accuracy shows the advantage of proposed Hough boxes. The average time of training our network for 52k, 2D images with 512$\times$512 size is about one day on single TitanX GPU with initially learning rate 0.0001. Instance segmentation for one 2D with 512$\times$512 is between 25-40 ms while it takes between 32-50 ms with MNC network.

## V. Conclusion

In this paper, we presented an automatic multitask learning network for tumor segmentation in medical images. We developed a new sequential convolutional neural network trained on the entire volume of medical images. We implemented a new bounding box detection technique and shared the code for research fraternity. Based on the promising segmentation results on two medical imaging benchmarks, the proposed architecture is robust against different sizes and shapes of tumor, different medical imaging protocols as well as various pathologies. Regarding the good accuracy and fast processing speed we think it has potential to be practically utilized for the routine clinical task. In future research, we plan to implement and investigate the potential of our architecture along discriminator network and train the whole architecture in adversarial way.

[2]https://github.com/MedInstSeg/

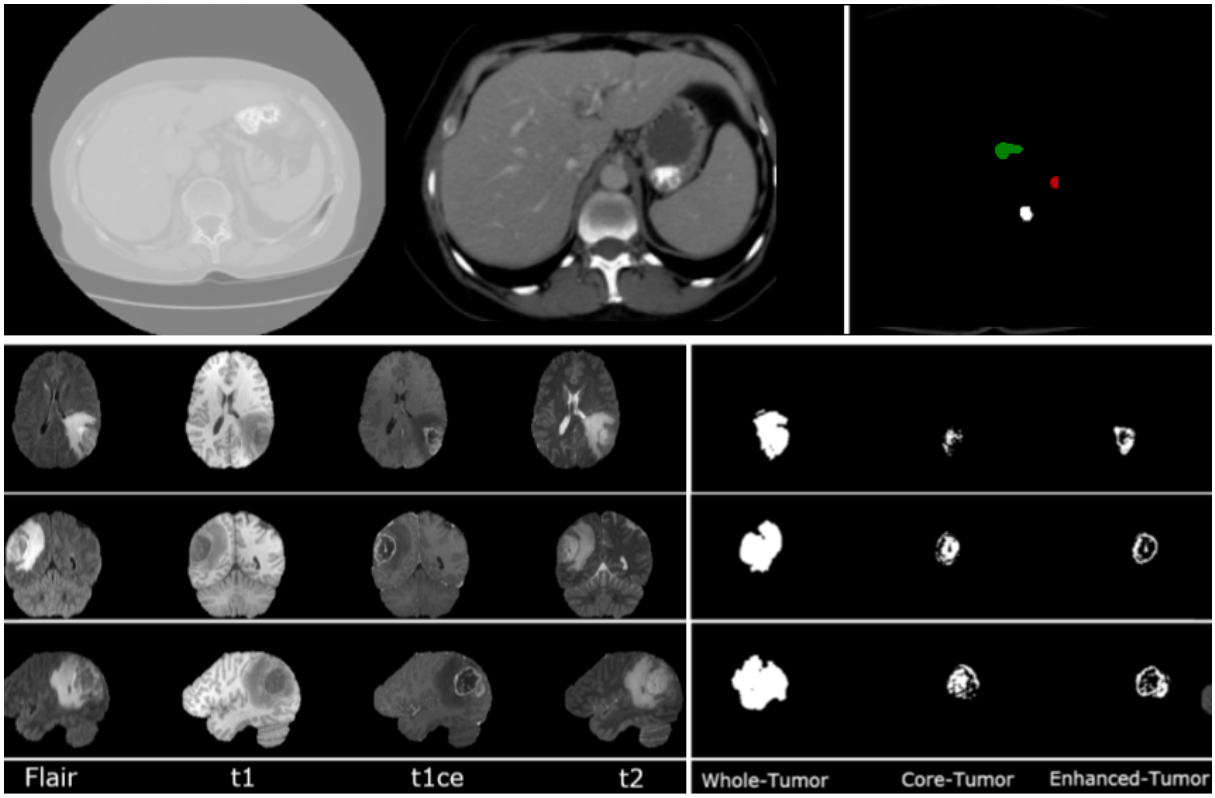


Fig. 4. The first row, first column shows the raw CT images, second column is final slice after Hounsfield unit windowing, and the last column is ground truth of tumor instance mask ready for the training process. Second row, first column shows MRI brain from BraTS-2016 in different acquisition plane and in four different modalities of Flair, t1, t1ce, and t2. The second column in second row is segmentation mask of whole brain, core tumor and enhanced tumor region which annotated by medical expert.

TABLE II

COMPARISON OF ACHIEVED ACCURACY IN TERM OF DICE METRICS, HAUSDORFF DISTANCE, SENSITIVITY, SPECIFICITY ON LITS BENCHMARK WITH MNC [18] AND OTHER APPROACHES. THE SECOND ROW SHOWS THE MNC RESULTS FROM SAME DATA AND SAME CONFIGURATION AS OURS WHERE SELECTED $10k$ RANDOMLY POINTS AS CENTER OF BOXES. WE DON'T APPLY PATIENT-WISE BATCH NORMALIZATION ON THE MNC EXPERIMENT.

| Liver lesion Seg. | Dataset | Dice | Hausdorff distance | Sensitivity | Specificity |
|---|---|---|---|---|---|
| our method | LiTS | 0.734 | 13.86 | 0.821 | 0.987 |
| MNC [18] | LiTS | 0.721 | 19.07 | 0.79 | 0.914 |
| Cascade-Unet + 3D CRF [6] | LiTS | 0.53 | 14.15 | 0.70 | 0.98 |
| Unet | LiTS | 0.70 | 18.06 | 0.84 | 0.93 |
| Cascade-Unet [6] | 3Dircadb | 0.56 | 14.15 | - | - |

TABLE III

COMPARISON OF ACHIEVED ACCURACY FOR WHOLE TUMOR REGION SEGMENTATION IN TERM OF DICE METRICS, HAUSDORFF DISTANCE, SENSITIVITY, SPECIFICITY ON BRATS2016 CHALLENGES WITH MNC [18] AND OTHER RELATED STATE-OF-THE-ART APPROACHES.

| Brain tumor Seg. | Dice | Hausdorff | Sensitivity | Specificity |
|---|---|---|---|---|
| our method | 0.823 | 12.04 | 0.89 | 0.98 |
| MNC [18] | 0.810 | 16.23 | 0.87 | 0.91 |
| InputCascade [17] | 0.85 | 14.15 | 0.80 | 0.91 |
| Unet | 0.80 | 22.53 | 0.82 | 0.87 |
| 3dUnet [11] | 0.81 | 13.88 | 0.82 | 0.96 |
| FCN | 0.80 | 13.65 | 0.84 | 0.96 |

TABLE IV

COMPARISON OF OUR IMPLEMENTED ARCHITECTURES IN SEMANTIC/ INSTANCE TUMOR SEGMENTATION WITH THE STATE-OF-THE-ART APPROACHES IN THE BRATS2016 AND LITS2016 BENCHMARK. IN THE THIRD AND FORTH COLUMN SHOW THE DETAIL OF ACHIEVED ACCURACY IN TERM OF DICE WITH DETAIL ON DETECTION AND SEGMENTATION STAGES.

| Brain tumor Seg. | Benchmark | Dice-Det | Dice-Seg | Sens | Spec |
|---|---|---|---|---|---|
| Proposed method | BraTS16 | 0.93 | 0.82 | 0.89 | 0.98 |
| MNC | BraTS16 | 0.90 | 0.81 | 0.87 | 0.91 |
| Unet | BraTS16 | - | 0.80 | 0.82 | 0.87 |
| Proposed method | LiTS17 | 0.78 | 0.73 | 0.82 | 0.98 |
| MNC | LiTS17 | 0.74 | 0.72 | 0.79 | 0.91 |
| Unet | LiTS17 | - | 0.70 | 0.84 | 0.93 |

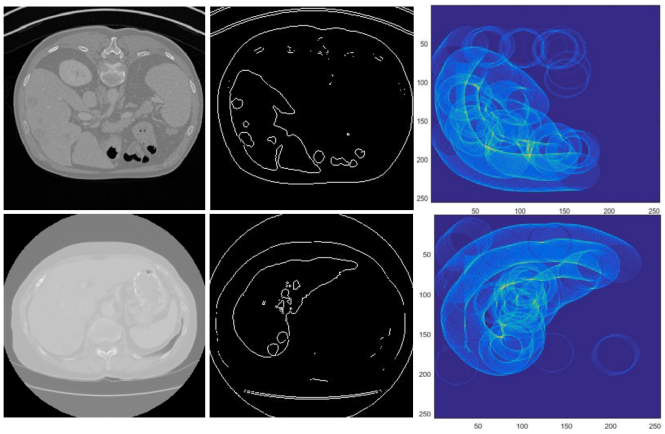

Fig. 5. The first column shows slice of CT images from LiTS dataset and second column shows the binary detected edge and last column is Hough transform accumulator for detection of curved shape.

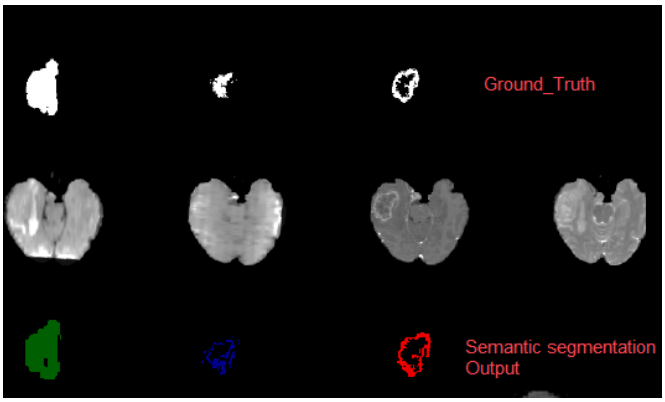


Fig. 6. The first row shows respectively the ground truth for whole tumor, core of tumor and enhance tumor. Second row is MRI image modalities provided by BraTS-2016. Third row shows semantic segmentation output of proposed method from left to right respectively for whole tumor, core of tumor and enhance tumor. The last row is visualization of our method with exact label of HGG/LGG.

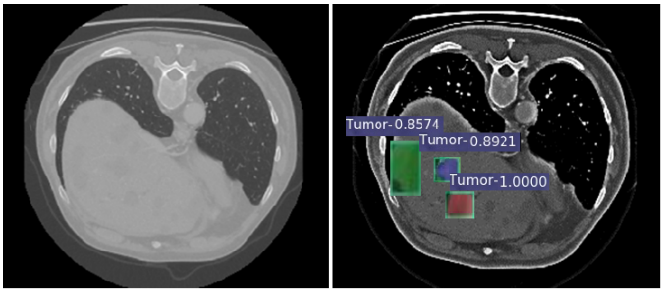


Fig. 7. Instance tumor segmentation result by proposed method on the 2D slice of abdomen CT of liver. The left side is the input image from test set and instance segmented boxes of tumors is shown on the right side.

TABLE V
BRAIN TUMOR SEMANTIC SEGMENTATION RESULT ON UNSEEN DATA WHICH WT, ET AND CT RESPECTIVELY IS ABBREVIATION OF WHOLE TUMOR, ENHANCED TUMOR AND CORE OF TUMOR.

| BraTS 2016 | Dice-WT | Dice-ET | Dice-TC |
|---|---|---|---|
| The proposed method | 0.82 | 0.69 | 0.77 |
| Havaei [17] | 0.82 | 0.90 | 0.81 |
| Tustison [28] | 0.75 | 0.55 | 0.52 |
| Fausto [20] | 0.62 | 0.50 | 0.61 |

## REFERENCES


[1] Rudy Bonavia Inda, Maria-del-Mar and Joan Seoane, "Glioblastoma multiforme: A look inside its heterogeneous nature.," in *Cancer Archive 226-239*, 2014.

[2] Mina Rezaei, Haojin Yang, and Christoph Meinel, "Deep neural network with l2-norm unit for brain lesions detection," in *Neural Information Processing - 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part IV*, 2017, pp. 798–807.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[5] Mina Rezaei, Haojin Yang, and Christoph Meinel, "Whole heart and great vessel segmentation with context-aware of generative adversarial networks," in *Bildverarbeitung für die Medizin 2018 - Algorithmen - Systeme - Anwendungen. Proceedings des Workshops vom 11. bis 13. März 2018 in Erlangen*, 2018, pp. 353–358.

[6] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A. Elshaer, Jana Lipková, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Felix Hofmann, Melvin D'Anastasi, Seyed-Ahmad Ahmadi, Georgios Kaissis, Julian Holch, Wieland H. Sommer, Rickmer Braren, Volker Heinemann, and Bjoern H. Menze, "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," *CoRR*, vol. abs/1702.05970, 2017.

[7] 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*. Springer International Publishing, 2015, pp. 234–241.

[8] Mina Rezaei, Konstantin Harmuth, Willi Gierke, Thomas Kellermeier, Martin Fischer, Haojin Yang, and Christoph Meinel, "A conditional adversarial network for semantic segmentation of brain tumor," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries - Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers*, 2017, pp. 241–252.

[9] Rich Caruana, *Multitask Learning*, Springer US, 1998.

[10] BH. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, MA. Weber, T. Arbel, BB. Avants, N. Ayache, P. Buendia, DL. Collins, N. Cordier, JJ. Corso, A. Criminisi, T. Das, H. Delingette, . Demiralp, CR. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, DL. Guo, A. Hamamci, KM. Iftekharuddin, R. Jena, NM. John, E. Konukoglu, D. Lashkari, JA. Mariz, R. Meier, S. Pereira, D. Precup, SJ. Price, TR. Raviv, SM. Reza, S. Ryan, D. Sarikaya, L. Schwartz, HC. Shin, J. Shotton, CA. Silva, N. Sousa, NK. Subbanna, G. Szekely, TJ. Taylor, OM. Thomas, NJ. Tustison, G. Unal, F. Vasseur, M. Wintermark, DH. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[11] P. H. A et al. Amorim, "3d unets for brain tumor segmentation in miccai 2017 brats challenge," pp. 9–14, 2017.

[12] Xiao Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239*, 2017.

[13] Andrew Beers1 et al., "Sequential 3d unets for brain tumor segmentation," pp. 20–28, 2017.

[14] Eunbyung Park and Alexander C Berg, "Learning to decompose for object detection and instance segmentation," *arXiv preprint arXiv:1511.06449*, 2015.

[15] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr, "Conditional random fields as recurrent neural networks," in *Proceed-*

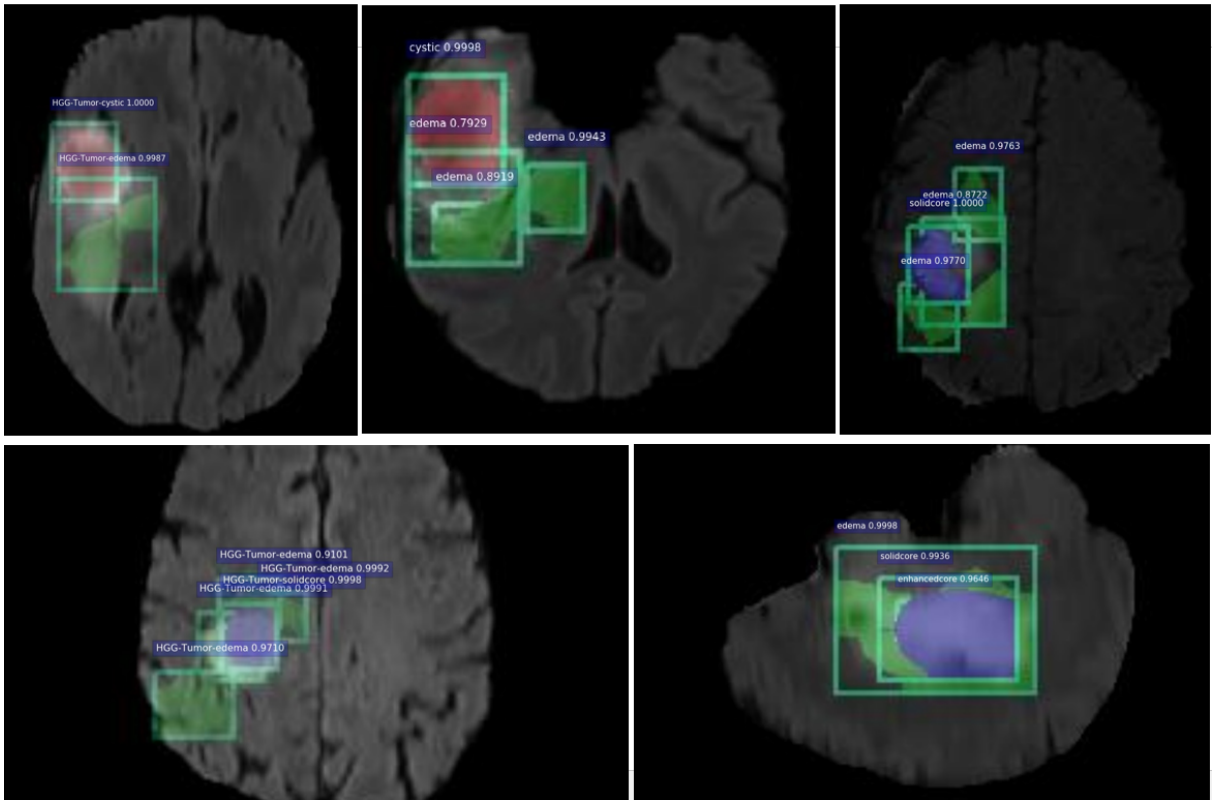


Fig. 8. The first row shows respectively the ground truth for whole tumor, core of tumor and enhance tumor. Second row is MRI image modalities provided by BraTS-2016. Third row shows semantic segmentation output of proposed method from left to right respectively for whole tumor, core of tumor and enhance tumor. The last row is visualization of our method with exact label of HGG/LGG.

*ings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[16] Jianxu Chen, Lin Yang, Yizhe Zhang, Mark S. Alber, and Danny Z. Chen, "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," *CoRR*, vol. abs/1609.01006, 2016.

[17] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[18] Jifeng Dai, Kaiming He, and Jian Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, 2016.

[19] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.

[20] Fausto Milletari et al., "Hough-cnn: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *CoRR*, 2016.

[21] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.

[22] Nathan Srebro and Adi Shraibman, "Rank, trace-norm and max-norm.," Springer.

[23] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.

[24] Dana H Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[25] John Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, , no. 6, pp. 679–698, 1986.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision–ECCV 2014*, 2014.

[27] Pedro O Pinheiro, Ronan Collobert, and Piotr Dollar, "Learning to segment object candidates," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., pp. 1990–1998. Curran Associates, Inc., 2015.

[28] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee, "N4itk: improved n3 bias correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[29] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.