International Journal of Machine Learning and Cybernetics (2022) 13:2435–2445
https://doi.org/10.1007/s13042-022-01536-4

ORIGINAL ARTICLE

Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net

Shidong Li[1] · **Jianwei Liu**[2,3] · **Zhanjie Song**[2,3]

Received: 10 June 2021 / Accepted: 24 February 2022 / Published online: 31 March 2022

Abstract

Since magnetic resonance imaging (MRI) has superior soft tissue contrast, contouring (brain) tumor accurately by MRI images is essential in medical image processing. Segmenting tumor accurately is immensely challenging, since tumor and normal tissues are often inextricably intertwined in the brain. It is also extremely time consuming manually. Late deep learning techniques start to show reasonable success in brain tumor segmentation automatically. The purpose of this study is to develop a new region-of-interest-aided (ROI-aided) deep learning technique for automatic brain tumor MRI segmentation. The method consists of two major steps. Step one is to use a 2D network with U-Net architecture to localize the tumor ROI, which is to reduce the impact of normal tissue's disturbance. Then a 3D U-Net is performed in step 2 for tumor segmentation within identified ROI. The proposed method is validated on MICCAI BraTS 2015 Challenge with 220 high Gliomas grade (HGG) and 54 low Gliomas grade (LGG) patients' data. The Dice similarity coefficient and the Hausdorff distance between the manual tumor contour and that segmented by the proposed method are 0.876 $\pm$0.068 and 3.594$\pm$1.347 mm, respectively. These numbers are indications that our proposed method is an effective ROI-aided deep learning strategy for brain MRI tumor segmentation, and a valid and useful tool in medical image processing.

Keywords Magnetic resonance imaging · Brain tumor segmentation · Deep learning · U-Net

1 Introduction

In brain tumor research, studying incidence rates, new cases, mortality, and survival rates etc., are becoming increasingly important [1]. Glioblastoma is the most common form of brain cancers and is highly aggressive malignancies with poor prognosis [2]. In the US, according to a report from 2011 to 2015, the 5-year and 10-year relative survival rates of patients [3] with malignant tumors were 35.0% and 29.3%, respectively. The annual average age-adjusted incidence rate of tumors was 23.03/10 million and the median survival time [3] of the patients was approximately 12–15 months, so the diagnosis of tumor shape and grade is critical to patient's treatment plan and prognosis. Multi-modality magnetic resonance imaging (MRI) is the principal scan tool of detection and diagnosis for glioblastoma tumor, since MRI provides superior contrast for soft tissues [4]. Magnetic resonance imaging (MRI) can identify, locate and classify brain tumors, capture different tumor phenotype spectra. In addition, multi-modality MRI can also capture the global information of the tumor. Accurate tumor contouring by using multi-modality MRI to identify features such as location, size, and invasion area plays a key role in computer-aided diagnosis [5].

Currently existing brain tumor segmentation techniques [6–9] can be briefly summarized as follows:

Threshold methods: these methods use local and global threshold to contour the tumor [10, 11]. However, the accuracy of segmentation is affected by the quality of MRI inhomogeneity correction. In addition, the optimal threshold is often difficult to identify.

✉ Jianwei Liu
jianweiliu@tju.edu.cn

[1] Department of Mathematics, San Francisco University, San Francisco, CA 94132, USA

[2] School of Mathematics, Tianjin University, Tianjin 300354, China

[3] Tianjin Key Laboratory of Brain-Inspired Intelligence Technology, Tianjin 300072, China

Edge-based methods: these methods detect tumor boundaries by the change of intensities in the MRI image [12, 13]. Generally, these methods are sensitive on the selection of thresholds among different intensities. In addition, the ambiguous and abnormal tumor boundaries can raise difficulties for these methods.

Region grow method: these methods delineate the tumor region by extracting sub-regions with similar neighboring voxels [14, 15]. However, these methods rely on the choice of initial seed point, which is a challenging task. In addition, these methods may be disturbed by MRI noise.

Watershed method: these methods segment the tumor contour by adding markers on landscape [16, 17]. The main challenge of these methods is that they may over-segment the tumor contour.

Shape-based methods: these methods generate the tumor contour of a new MRI sequence by comparing it with other template sequence, such as using atlas and deformation [18]. However, the accuracy of atlas reconstruction and deformable registration will affect the performance of these methods.

Clustering methods: these methods extract features from MRI and use unsupervised and supervised classification model to assign MRI voxels into different groups. The unsupervised model, such as fuzzy c-means [19], k-means [20], and principal component analysis (PCA) [21, 22], do not need a training procedure. The supervised model, such as random forest (RF) [23], markov random field [24], support vector machine (SVM) [25], extreme learning machine [26] and deep learning [7], need to first train a classification or segmentation model, then feed a new acquired MRI into the trained network to obtain a segmented contour of the brain image. A late example is by Yang et al. where PCA feature extraction is conducted, they are able to diagnose breast tumors by using SVM with differential evolution-based parameter tuning [22].

Compared with deep learning methods, the performance of other clustering methods rely on how well the handcraft features can represent the brain tumor in MRI images, this is a huge challenge since tumor shapes vary dramatically among different types. Irregular tumor tissue will seriously affect the accuracy of the classification. In addition, since these methods need to extract features of multiple levels from MRI images followed by smaller patch or single voxel assignment, the computational cost of these methods is expensive. In the meantime, a number of useful classification methods have been developed recently. For example, Luo et al. developed a automatic left ventricle segmentation method which combined a layered extreme learning machine with a new location method [26], the training cost reduced while the segmentation accuracy improved. Liu et al. proposed a new method for feature selection to classify benign and malignant tumors with a new twin support vector machine [27].

In the meantime, recent deep learning methods for semantic segmentation, especially the end-to-end fully convolutional neural networks (FCNN), have shown great performance in medical image segmentation [28]. Zhang et al. studied FCNN with adjusted network architecture for brain tumor segmentation [29]. Ben et al. developed a fused segmentation technique by combining results of a 2D and 3D incremental network [30]. The incremental network is a particular type of FCNNs. In [31], a technique applying densely connected CNN for brain MRI tumor segmentation was studied. Chang Liu et al. used the deep neural networks to automatically segment the prostate of CT images [32], and got good results. A new deep neural network (P-DNN) is proposed in [33] to extract complementary features from MR prostate images, and outperforms traditional deep neural networks in terms of results. Ryo Ito et al. segmented brain tissue by semi-supervised deep learning method [34], which leverages the image registration technology.

Although these methods produce relatively effective performances on brain tumor segmentation, the accuracy of these methods may be limited by GPU memory capacities. Because, for 3D whole volume images and corresponding manual contouring, the training data is often too huge to train a segmentation model on a typical GPU. A possible solution is to extract smaller 3D patches with a sliding window as input. However, this strategy will lose global and spatial information, and thus produce a relatively inaccurate segmentation.

To resolve this issue, this study aims at developing a new region-of-interest-aided (ROI-aided) deep learning method to exclude the effect of irrelevant tissues in brain MRI images, and thus enhances the accuracy of segmentation.

Recently, Feng et al. proposed a cropped 3D image-based deep CNN method for their thoracic organs-at-risk multi-label segmentation [35]. They used 3D U-Net to localize each organ, and the images which only contain one organ were served as the input to each individual organ segmentation network. The final segmentation results were obtained by merging the segmentation maps of each organ. The method they proposed earned the second place in the live phase of the 2017 AAPM Thoracic Auto-segmentation Challenge. In the subsequent ongoing phase, they used a newly developed testing augmentation approach and earned the first place [35]. Our study is inspired by their work.

The strategies of this work include the following.

First, as brain tumors tend to group in a location, a 2D U-Net [36] is used to localize a ROI of tumor by exploiting the whole slices' information and then we obtain the ROI of tumor from a coarse segmentation result. Other than down sample the image as in [35], MRI images of original resolution are used, the aim is to identify the specific location

of ROI, which is important for our next step of accurate segmentation.

Second, a 3D U-Net [36] is used to acquire a precise segmentation within the specified ROI by exploiting spatial and structural information from the 3D volume of ROI.

The rest of this article is organized as follows. In Sect. 2, an overview of the proposed brain tumor segmentation framework with the description of ROI-aided strategy is provided, followed by detailed descriptions of image acquisition, 2D and 3D U-Net architectures, evaluating metrics, and comparing algorithms. In Sect. 3, the proposed method is compared with other algorithms for performances. Extended discussions about the proposed technique and future studies are provided in Sect. 4.

2 Materials and methods

2.1 Overview and ROI-aided strategy

Any learning based segmentation technique typically consists of two phases, namely training and testing. Figure 1 outlines the schematic flow chart of the proposed method. In the training phase (upper part of Fig. 1), a 2D U-Net localization model is trained. The multi-modal MRI sequences of each training patient are used as input images and the associated patient's manual brain tumor contour is regarded as the learning target. Slices of images extracted from multi-modal brain MRI images, i.e., Flair, T1, T1-contrast-enhanced (abbreviated as T1c in Fig. 1), and T2 scan, respectively, are concatenated and used as multi-channel inputs. The corresponding slice of binary mask of manual contour is used as ground truth to supervise the training of the 2D U-Net. The input image size of this network is set by the size of slice in axial plane, which is 240×240 (voxel size). The output of 2D U-Net is a coarsely generated binary mask of tumor, which is termed as localization by this study.

After generating the coarse binary mask of tumor via 2D U-Net, an ROI is identified by first locating the mass center of the mask and then a 3D bounding box centered at that center is created to fully enclose the tumor. Considering that the maximum height, width, and depth of tumors on MICCAI BraTS 2015 Challenge data [5] are 162, 125, and 111 for high Gliomas grade (HGG) data and 141, 112, and 89 for low Gliomas grade (LGG) data, respectively, oversize window size will impact the training efficiency and undersize window size will lead to insufficient extraction, so the window size is set to 128×128×128 pixels. To cover the tumor, if one ROI cannot sufficiently enclose the generated coarse binary mask of tumor, neighboring ROIs of the same size are also extracted. The correlated coordinates of the ROI/ROIs are recorded so that the segmented binary masks

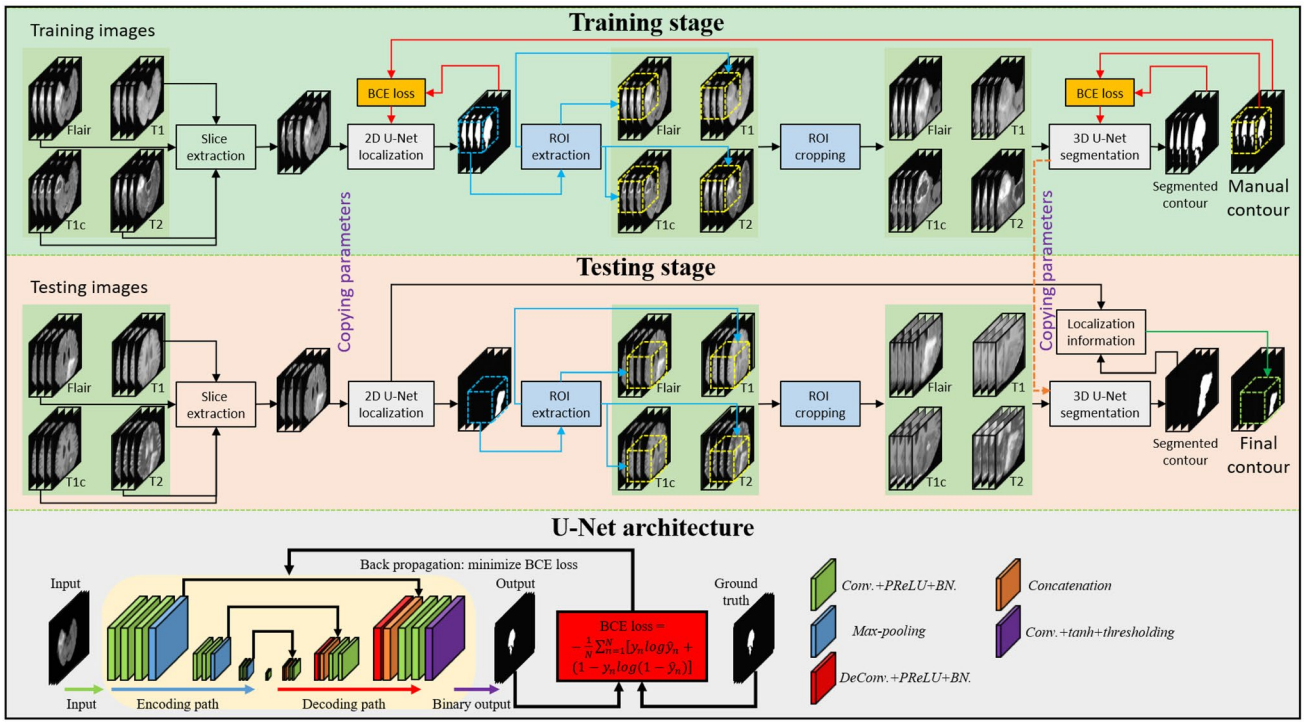

Fig. 1 The schematic flow diagram of the proposed method. The upper part shows training procedure of 2D localization and 3D segmentation models. The middle part shows the testing stage for a new data. The lower part shows the architecture of U-Net used for both 2D localization and 3D segmentation models

within ROI/ROIs can be re-located back to the original volumetric space.

After that, we crop the MRI images within the ROI, and feed the cropped image to the 3D U-Net to derive the binary mask within ROI. The cropped MRI images from four scans are inputs, and the manual contour of corresponding ROI is set as target to train the 3D U-Net for accurate brain tumor segmentation within the ROI.

Evidently, the ROI strategy is to exclude the effect of irrelevant tissues in brain MRI images, and enhances the accuracy of segmentation. Different from [35] where down sampled images were used for cropping, MRI images of original resolution are used in the proposed ROI approach. Compared to manual cropping, the ROI method uses 2D U-Net network to locate the brain tumor tissues, followed by a 3D window to fully enclose the tumor tissues to implement the cropping.

In the testing/segmentation phase (middle part of Fig. 1), for a new multi-modal MRI data, the location of tumor ROI and the tumor segmentation within that ROI are obtained following the same sequence as in training. The final tumor segmented contour is obtained by locating the segmented tumor within ROI in the entire 3D image.

2.2 Image acquisition

In this study, we collect the brain MRI data and corresponding manual tumor contour from a benchmark public database, i.e., the multi-modal brain tumor image segmentation challenge held in conjunction with MICCAI 2015 (MICCAI-BRATS 2015). This is a large database of brain tumor MRI scans in which the relevant tumor structures have been well delineated by experienced physicians [5]. The dataset include multimodality MRI and the brain tumor image that were delineated by both clinicians and radiologists. The MRI images (acquired at four centers, i.e., Heidelberg University, Massachusetts General Hospital, Debrecen University, and Bern University) were scanned by MRI machines of different vendors with 1.5T and 3T field strengths [5]. The multi-modal MRI scans of 220 HGG and 54 LGG patients who have received pre- and post-therapy are used for training and testing in this study. The maximum height, width, and depth of tumors on MICCAI BraTS 2015 Challenge data [5] are 162, 125, and 111 for high Gliomas grade (HGG) data and 141, 112, and 89 for low Gliomas grade (LGG) data, respectively.

The image data set share the following four MRI comparisons [5] and all images are skull stripped.

(1) FLAIR: T2-weighted FLAIR image, 2–6 mm slice thickness, sagittal, axial, or coronal 2D acquisitions.
(2) T1: T1-weighted, native image, 1–6 mm slice thickness, axial or sagittal 2D acquisitions.
(3) T1c: T1-weighted, contrast-enhanced image, 1 mm isotropic voxel size, 3D acquisition.
(4) T2: T2-weighted image, 2–6 mm slice thickness, axial 2D acquisition.

Volumes of each patient's MRI images are rigidly registered to its T1c MRI scan, which has the highest spatial resolution over all scans, and are then restored to the same resolution by resampling ($1\times1\times1$ mm^3) in that challenge data [5] to achieve the homogeneity of these data. The voxel size of these patients' MRI data is $240\times240\times155$ mm^3. To unify these data, we use a normalization strategy to normalize all MRI intensities to the range [– 1, 1]. The normalization is performed as follows:

$$I_{normalized} = \frac{I - \mu(I)}{max_{95}(I) - min_{95}(I)} \quad (1)$$

where I denotes the original MRI images of a patient's MRI scan, $\mu(I)$ denotes the mean value, $max_{95}(I)$ denotes the 95% maximum intensity value, $min_{95}(I)$ denotes the 95% minimum intensity value.

The quality of MRI images may be affected by image inhomogeneity issue. To resolve this problem, a N4 bias field correction method, which is a universal method for correcting non-uniformity of low frequency intensity, is used in the SimpleITK library of Python package.

2.3 U-Net architecture

For both 2D localization and 3D segmentation models, the binary mask of manual contour is used as the learning target of multi-modal MRI image for our proposed method. In order to fully exploit the multi-scale information, including spatial and structural information, U-Net architecture is used for both 2D and 3D models. This is because that the U-Net system can extract not only low-frequency feature maps from previous convolution layers but also high-frequency feature maps from subsequent convolution layers. These multi-level feature maps may be helpful for representing the tumor boundaries in multiple levels.

As shown in the lower part of Fig. 1, the U-Net architecture is divided into two parts: encoding path and decoding path. In the encoding path, the input image slices (for 2D localization) and ROI volumes (for 3D segmentation) first go through three convolution operators with a max-pooling to reduce the size of feature maps. Each convolution operator is implemented by a convolution layer followed by a Parametric Rectified Linear Units (PReLU) and a batch normalization (shorted as Conv.+PReLU+BN in Fig. 1). Then the feature maps go through three deconvolution operators to

obtain an equal-sized output, which is called end-to-end in FCNN [28]. In order to preserve the multi-level features, a concatenation is performed to enlarge the number of feature maps between the outputs of two equal-sized feature maps obtained from encoding and decoding path, respectively. Finally, the feature maps are fed into a convolution layer to shorten the number of feature maps to only two, and followed by a "tanh" layer to polarize the feature maps. The two-channel feature maps are regarded as posterior of the tumor tissue and the normal tissue. After that, a threshold is used to obtain the binary mask of tumor.

Binary cross entropy (BCE) loss, which measures the dissimilarity between generated binary mask of tumor and the binary mask of manual contour, is used to supervise the model. Let $\{S_i : i \in I\}$ denote all values of voxels of generated binary mask of tumor, and $\{M_i : i \in I\}$ be values of voxels of binary mask of manual contour. The binary cross entropy is defined as follows:

$$Loss_{BCE} = -\sum_{i \in I} M_i log(S_i) - \sum_{i \in I} (1 - M_i) log(1 - S_i) \tag{2}$$

In order to enhance the robustness against the inaccuracies in both tumor localization and segmentation, data augmentation, including flipping, rotation, and elastic deformation, is used to expand the variety of training data.

2.4 Quantitative measurements

We use ten-fold cross-validation experiments to evaluate the proposed brain tumor segmentation algorithm. Specifically, following the principle of equal distribution, we randomly separate the patients' data (both 220 HGG and 54 LGG data) into ten groups, where the first nine groups have data from 27 patients, and the last group have data from 31 patients. For any of such ten-fold cross-validation experiments, one group of data set is taken as the validation data for performance evaluation, the rest nine groups are regarded as the training data. This experiment is repeatedly performed ten times, with each group served as validation data once.

To evaluate the performance, the measuring metrics are calculated by comparing the segmented brain tumor contour with the known manual contour by experts. Specifically, the Dice similarity coefficient (DSC), the Hausdorff distance (HD), and the mean surface distance (MSD) are used to evaluate the accuracy of the proposed method.

DSC is applied to measure the overlapping of the segmented contour obtained by proposed tumor segmentation method and the manual contour delineated from experienced physicians.

$$DSC = \frac{2 \cdot |S \bigcap M|}{|S| + |M|} \tag{3}$$

where S denotes the segmented contour, M denotes the corresponding manual contour, $|\cdot|$ denotes the number of voxels within that contour.

HD and MSD measure the surface area differentiation between the segmented and manual contour. HD is used to evaluate the maximum difference between the surface area of segmented contour S and corresponding manual contour M, MSD is applied to measure the average difference between the surface area of segmented contour S and corresponding manual contour M, given as

$$H(S, M) = max(h(S, M), h(M, S)) \tag{4}$$

$$MSD = \frac{h(S, M) + h(M, S)}{2} \tag{5}$$

where $h(S, M) = \max_{s \in S} \min_{m \in M} \|s - m\|$, $h(M, S) = \max_{m \in M} \min_{s \in S} \|m - s\|$, $\|\cdot\|$ denotes the Euclidean distance.

To prove that ROI-aided strategy can exclude the influence of irrelevant tissue in MRI scans, we compare the proposed method with 3D U-Net without using 2D U-Net for localization (called as pure 3D U-Net) and set the patch size as 128×128×128. In order to illustrate that the whole 3D ROI input can exploit more spatial information, we compare the proposed method with 2D U-Net, which regard the slice of transverse plane as input (called as pure 2D U-Net).

The latest well known V-Net [28] is used in the comparison studies. V-Net uses Dice-loss as the loss function to supervise the model and adjusts network architecture by adding more convolution layers and replacing the concatenation operator with element sum operator.

The proposed method is also compared with the recent densely connected FCN algorithm (called as DCFCN) [31]. The main idea of this algorithm is to utilize densely connected convolutional blocks to boost the performance.

To illustrate the significant improvement of the proposed method against competitive techniques, a paired two-tailed t-test is used between the numerical results calculated by the proposed method and that by the comparing approaches. All parameters among studied techniques are set by their best performances.

3 Results

3.1 Robustness of ROI-aided strategy

In the proposed method, we first use 2D U-Net to obtain a coarse binary mask of tumor and then set a ROI to cover and center this coarse binary mask. The center position of ROI is originally set as the mass center of the coarse binary mask. In order to ensure that this ROI location identification is robust for the brain tumor segmentation, we randomly

choose one group as validation data from the above ten groups of data in Sect. 2.4, and set all possible eight groups from the rest nine groups as training set. This procedure has altogether $C_9^8 = 9$ times of experiments. The average metrics of each experiment is used to compare the differences among various methods.

Table 1 shows the DSC, HD, and MSD of the segmentation results of these nine experiments. Among the nine experiments of the proposed method, the DSC, HD, and MSD do not show any significant change (with all the P-value between each two experiments larger than 0.05). Thus, it demonstrates that ROI location of the proposed method is not sensitive to different test.

3.2 Contribution of ROI-aided strategy

To demonstrate the efficacy of ROI-aided strategy, we compare the results of proposed method with pure 3D U-Net and pure 2D U-Net as previously mentioned. Figure 2 shows the visual results on axial view of these three methods. Sub-images (a1–a8) show an extreme case of a HGG patient, where the Flair image (a1) loses almost half of the soft tissue region as compared with the rest three scans (a2–a4). Sub-images (b1–b8) show a normal case of a LGG patient. As is shown in (b1–b8), with the regular boundary contrast appeared in the multi-modal MRI images, all three methods provide similar segmentation results, which are very close to the manual contour.

However, as shown in (a1–a8), when normal tissue intensities of T2 images are at similar levels of tumor intensities (as pointed out in (a4) by yellow arrow), the pure 3D U-Net introduces some false segmentation, whereas pure 2D U-Net and the proposed method show no false tumor boundary. Furthermore, the proposed method can recover more details of the tumor boundary than that pure 2D U-Net can, as shown in sub-image comparison of (a8) vs. (a7), we believe the ability of showing boundary details is due to the exploitation of features from 3D volume of the whole ROI. In the meantime, misclassification of pure 3D U-Net may be caused by patch-based feature extraction that causes the loss of some spatial information, this misclassified region should be ventricle tissue not the tumor tissue.

Table 2 is used to test the improvement of the ROI-aided strategy by using ten-fold cross-validation experiments. The results demonstrate that the proposed method significantly improves the performance compared with other methods on the DSC metric (P-value<0.05). On the HD metric comparison, the proposed method is significantly better than that of pure 3D U-Net, while no significant improvement over that of pure 2D U-Net. As for the MSD metric comparison, the three methods show comparable results.

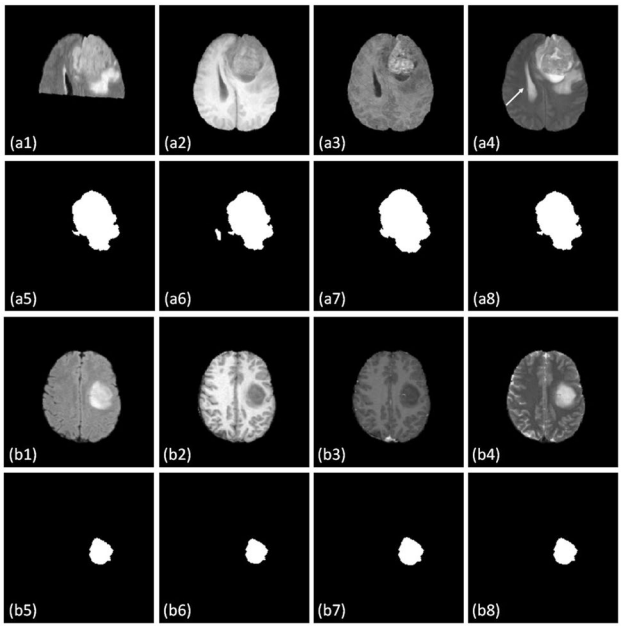

Fig. 2 The visual results of the proposed and comparing methods on transverse plane. a1–a8 Show a HGG patient's Flair, T1, T1c, T2 image, binary mask of manual contour, result obtained by pure 3D U-Net, result obtained by pure 2D U-Net, and the result obtained by the proposed method, respectively. b1–b8 Show a LGG patient's images as same image sequence with (a1–a8)

Table 1 Numerical results with different test

Experiment#	1	2	3	4	5	6	7	8	9
DSC									
Mean	0.879	0.881	0.877	0.884	0.875	0.876	0.877	0.868	0.886
Std	0.060	0.060	0.061	0.059	0.060	0.060	0.060	0.060	0.060
HD									
Mean	3.549	3.543	3.575	3.594	3.594	3.515	3.553	3.501	3.505
Std	1.380	1.426	1.357	1.391	1.373	1.414	1.347	1.381	1.397
MSD									
Mean	0.283	0.287	0.278	0.271	0.274	0.278	0.287	0.278	0.287
Std	0.070	0.070	0.071	0.070	0.070	0.070	0.070	0.070	0.070

Table 2 Numerical comparison between proposed method with pure 3D and 2D U-Net among all patients' data

Comparing method	Pure 3D U-Net	Pure 2D U-Net	Proposed
DSC			
Mean	0.858	0.860	0.877
Std	0.080	0.069	0.060
P-value	<0.001	0.039	N/A
HD			
Mean	5.102	3.638	3.553
Std	1.702	1.588	1.347
P-value	<0.001	0.051	N/A
MSD			
Mean	0.290	0.285	0.287
Std	0.070	0.070	0.070
P-value	0.764	0.893	N/A

The P-value of each column is calculated by comparing the results of proposed method and the results of the comparing method in that column

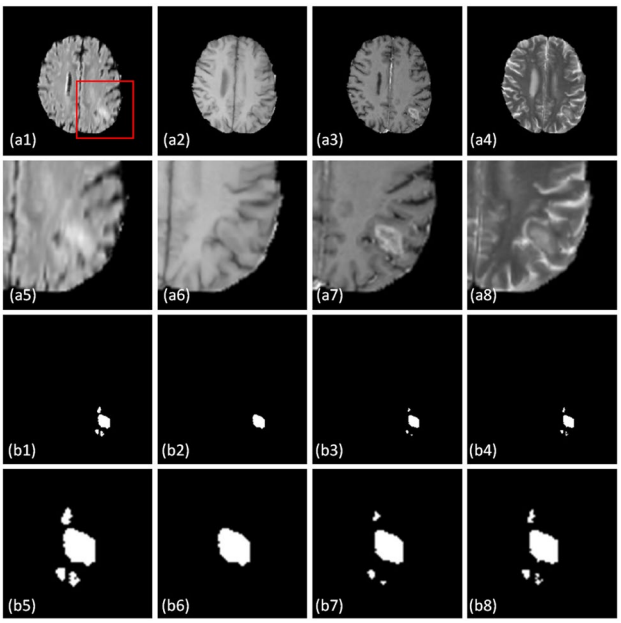

Fig. 3 The visual results of the proposed and other state-of-the-art methods on axial view. a1–a4 Show a patient's Flair, T1, T1c, and T2 images. a5–a6 Show the zoomed-in images of (a1–a4), the zoomed-in window location is shown as the red rectangle in (a1). b1–b4 Show the binary mask by the manual contour, by the V-Net, by the DCFCN, and by the proposed method, respectively. b5–b8 Show the zoomed-in images of (b1–b4) within the same zoomed-in window as shown in the red rectangle of (a1)

3.3 Comparison with state-of-the-art methods

To further validate the proposed method, its performance is compared against previously mentioned V-Net [28] and DCFCN [31]. Figure 3 shows the visual segmentation results of one patient. The aim of this figure is to compare the performance when the boundary contrast of tumor on MRI images are not very clear to distinguish. As is shown in sub-images (a1) and (a4), the tumor edges of the Flair and T2 images are smooth and ambiguous. In addition, the T1 image shows nearly minimal tumor structure in sub-image (a2). Under this extreme situation, our proposed method and DCFCN still have similar binary mask of the tumor to that of manual tumor contour, but the V-Net method cannot. Furthermore, the result of the proposed method is much more like the manual contour.

Table 3 compares the numerical results of the proposed method against two other techniques. It shows clearly that the proposed method outperforms the other two methods on DSC and HD metrics significantly.

For comparison, the numerical results from the BRATS-2015 leaderboard are carefully studied. The top-one ranked most recent segmentation model [37] for this brain tumor segmentation has the DSC performance of 0.87, the second to fourth top-ranked models [38–40] have the DSC performance of 0.85, and the fifth top-ranked model [41] has the DSC performance of 0.84.

In the meantime, the DSC performance of our proposed method (around 0.876) is close to top-one model and outperforms the rest four models, which demonstrates the efficiency of our proposed method.

4 Discussion

We propose a new brain tumor segmentation method utilizing the new ROI-aided strategy into the U-Net architecture to automatically locate the tumor and then segment the

Table 3 Numerical comparison between proposed method with V-Net and DCFCN among all patients' data

Comparing method	V-Net	DCFCN	Proposed
DSC			
Mean	0.859	0.867	0.876
Std	0.079	0.077	0.068
P-value	<0.001	0.042	N/A
HD			
Mean	5.222	3.926	3.594
Std	1.800	1.642	1.347
P-value	<0.001	0.035	N/A
MSD			
Mean	0.290	0.290	0.289
Std	0.069	0.070	0.069
P-value	0.884	0.893	N/A

The P-value of each column is calculated by comparing the results of proposed method and the results of the comparing method in that column

tumor contour. Our proposed method is significantly better compared with the state-of-the-art methods on DSC and HD metrics as described in Table 3. The application of the new ROI-aided strategy achieve significant improvement on the DSC metric as compared with pure 2D and 3D U-Net without using the new ROI-aided strategy.

Different from recent fully convolutional network (FCN)-based methods, which directly perform segmentation on whole input volume, our proposed network is a cascaded network including two U-Nets. The first U-Net provides a rough tumor segmentation, and derives the location and boundary of tumor, namely the region-of-interest (ROI) of the tumor. Then, the second U-Net performs the fine segmentation on the image that is cropped within the ROI of the tumor. The effectiveness of our proposed cascading U-Net method can be summarized in two folds. First, useless information from the irrelevant regions is excluded. These include, for example, the bony and brain stem structures, which have higher or equal tissue contrast as compared to tumor. These regions may introduce additional bias if one FCN is directly applied for tumor segmentation. This is the purpose of the first U-Net to obtain the ROI of the tumor, i.e., locating the tumor and its relative small region of interest.

Secondly, the challenge of segmenting tumor through deep learning is that tumor region is relatively small than that of the whole brain volume. Thus, if one takes the whole brain volume as input for a direct segmentation, the imbalance issue between positive labels (tumor voxels) and negative labels (non-tumor voxels) will be significant, which may cause convergence difficulties for recent deep learning-based models. To resolve this issue, we crop the image within ROI that is obtained via first U-Net and feed it to the second U-Net. The non-tumor and tumor regions within the ROI would be at a relatively equal level. The imbalance issue for the second U-Net would be diminished. Thus, the accuracy of the obtained fine segmentation of tumor would be enhanced as compared to recent FCN-based methods.

So, two U-Nets are used for brain tumor segmentation in our proposed method. The 2D U-Net aims to roughly determine the location (i.e., ROI) of brain tumor and the 3D U-Net aims to perform fine segmentation on the ROI. The details of our networks design are as follows.

Tables 4 and 5 show the structural parameters of 2D and 3D U-Nets, respectively. Adam gradient optimizer is used to optimize the two networks. The learning rate of adam gradient optimizer is set to 1.0×10^{-5} and the same-padding operation is used. The batch size of 2D U-Net training is set to 20 and the batch size of 3D U-Net training is set to 4. It takes about 0.5 h with 200 epochs during training the 2D U-Net, and takes about 7 h with 200 epochs during training the 3D U-Net. For a new input MRI, the 2D U-Net takes about 0.5 min to obtain the brain tumor location, and the 3D U-Net takes about 2 min to derive the tumor segmentation.

Table 4 The architecture of 2D U-Net

Backbone	Filter and stride	Feature	Input shape
Input	N/A	N/A	$240 \times 240 \times 4$
Conv_0_0	$3 \times 3\&1 \times 1$	16	$240 \times 240 \times 4$
Conv_0_1	$3 \times 3\&1 \times 1$	16	$240 \times 240 \times 16$
Conv_0_2	$3 \times 3\&1 \times 1$	16	$240 \times 240 \times 16$
MaxPool_0	$2 \times 2\&2 \times 2$	–	$240 \times 240 \times 16$
Conv_1_0	$3 \times 3\&1 \times 1$	32	$120 \times 120 \times 16$
Conv_1_1	$3 \times 3\&1 \times 1$	32	$120 \times 120 \times 32$
MaxPool_1	$2 \times 2\&2 \times 2$	–	$120 \times 120 \times 32$
Conv_2_0	$3 \times 3\&1 \times 1$	64	$60 \times 60 \times 32$
Conv_2_1	$3 \times 3\&1 \times 1$	64	$60 \times 60 \times 64$
MaxPool_2	$2 \times 2\&2 \times 2$	–	$60 \times 60 \times 64$
Conv_3_0	$3 \times 3\&1 \times 1$	128	$30 \times 30 \times 64$
Conv_3_1	$3 \times 3\&1 \times 1$	128	$30 \times 30 \times 128$
DeConv_2	$3 \times 3\&2 \times 2$	64	$30 \times 30 \times 128$
Concate_2	–	–	$60 \times 60 \times 64$
Conv_2_2	$3 \times 3\&1 \times 1$	64	$60 \times 60 \times 128$
Conv_2_3	$3 \times 3\&1 \times 1$	64	$60 \times 60 \times 64$
DeConv_1	$3 \times 3\&2 \times 2$	32	$60 \times 60 \times 64$
Concate_1	–	–	$120 \times 120 \times 32$
Conv_1_2	$3 \times 3\&1 \times 1$	32	$120 \times 120 \times 64$
Conv_1_3	$3 \times 3\&1 \times 1$	32	$120 \times 120 \times 32$
DeConv_0	$3 \times 3\&2 \times 2$	16	$120 \times 120 \times 32$
Concate_0	–	–	$240 \times 240 \times 16$
Conv_0_2	$3 \times 3\&1 \times 1$	16	$240 \times 240 \times 32$
Conv_0_3	$3 \times 3\&1 \times 1$	16	$240 \times 240 \times 16$
Conv_final	$3 \times 3\&1 \times 1$	2	$240 \times 240 \times 16$

We then locate the segmentation back to the original image coordinate according to the localization obtained via 2D U-Net. This procedure takes within 1 s.

Our method can perform segmentation within 1 s for a new MRI image from a newly arrived patient, which will be an efficient tool for the brain tumor contouring and a potential tool for solving the labor-consuming issue of manual contour during clinic. As is shown in Table 3 and the comparison with most recent brain tumor segmentation methods listed in leaderboard of BRATS2015, the DSC performance is better than most of these methods and is comparable with top one ranked method, which demonstrates the efficiency and accuracy of our proposed method. In total, our proposed method takes about 7 GB during training. The proposed method can be also used in segmenting of tumor and lesion of different body sites and different image modality. For example, a recent study [42] of lung CT infection segmentation for COVID-19 pneumonia patient used the strategy of cascaded network. In future, we will evaluate the performance of proposed method on this dataset.

Our method is suitable for segmentation on different medical image modalities. However, the intensity values

Table 5 The architecture of 3D U-Net

Backbone	Filter and stride	Feature	Input shape
Input	N/A	N/A	$128 \times 128 \times 128 \times 4$
Conv_0_0	$3 \times 3 \times 3\&1 \times 1 \times 1$	16	$128 \times 128 \times 128 \times 4$
Conv_0_1	$3 \times 3 \times 3\&1 \times 1 \times 1$	16	$128 \times 128 \times 128 \times 16$
Conv_0_2	$3 \times 3 \times 3\&1 \times 1 \times 1$	16	$128 \times 128 \times 128 \times 16$
MaxPool_0	$2 \times 2 \times 2\&2 \times 2 \times 2$	–	$128 \times 128 \times 128 \times 16$
Conv_1_0	$3 \times 3 \times 3\&1 \times 1 \times 1$	32	$64 \times 64 \times 64 \times 16$
Conv_1_1	$3 \times 3 \times 3\&1 \times 1 \times 1$	32	$64 \times 64 \times 64 \times 32$
MaxPool_1	$2 \times 2 \times 2\&2 \times 2 \times 2$	–	$64 \times 64 \times 64 \times 32$
Conv_2_0	$3 \times 3 \times 3\&1 \times 1 \times 1$	64	$32 \times 32 \times 32 \times 32$
Conv_2_1	$3 \times 3 \times 3\&1 \times 1 \times 1$	64	$32 \times 32 \times 32 \times 64$
MaxPool_2	$2 \times 2 \times 2\&2 \times 2 \times 2$	–	$32 \times 32 \times 32 \times 64$
Conv_3_0	$3 \times 3 \times 3\&1 \times 1 \times 1$	128	$16 \times 16 \times 16 \times 64$
Conv_3_1	$3 \times 3 \times 3\&1 \times 1 \times 1$	128	$16 \times 16 \times 16 \times 128$
DeConv_2	$3 \times 3 \times 3\&2 \times 2 \times 2$	64	$16 \times 16 \times 16 \times 128$
Concate_2	–	–	$32 \times 32 \times 32 \times 64$
Conv_2_2	$3 \times 3 \times 3\&1 \times 1 \times 1$	64	$32 \times 32 \times 32 \times 128$
Conv_2_3	$3 \times 3 \times 3\&1 \times 1 \times 1$	64	$32 \times 32 \times 32 \times 64$
DeConv_1	$3 \times 3 \times 3\&2 \times 2 \times 2$	32	$32 \times 32 \times 32 \times 64$
Concate_1	–	–	$64 \times 64 \times 64 \times 32$
Conv_1_2	$3 \times 3 \times 3\&1 \times 1 \times 1$	32	$64 \times 64 \times 64 \times 64$
Conv_1_3	$3 \times 3 \times 3\&1 \times 1 \times 1$	32	$64 \times 64 \times 64 \times 32$
DeConv_0	$3 \times 3 \times 3\&2 \times 2 \times 2$	16	$64 \times 64 \times 64 \times 32$
Concate_0	–	–	$128 \times 128 \times 128 \times 16$
Conv_0_2	$3 \times 3 \times 3\&1 \times 1 \times 1$	16	$128 \times 128 \times 128 \times 32$
Conv_0_3	$3 \times 3 \times 3\&1 \times 1 \times 1$	16	$128 \times 128 \times 128 \times 16$
Conv_final	$3 \times 3 \times 3\&1 \times 1 \times 1$	2	$128 \times 128 \times 128 \times 16$

of different image modalities, such as CT, cone-beam CT (CBCT), and ultrasound image, will be different. As a result, different normalizations for these datasets should be applied. In addition, it should be noticed that the image quality of different image modalities is affected by different bias issue. For example, in our task, the MRI is affected by inhomogeneity issue, thus we use an inhomogeneity correction tool as preprocessing step. Scatter artifact and ring artifact will affect the image quality of CBCT and CT, so scatter correction [43] and ring artifact reduction [44] should be performed before training and testing the model. In ultrasound, speckle noise may affect the segmentation performance, so speckle noise reduction method [45] should be performed.

There are still more practical tasks to be studied. First of all, since the MRI images of MICCAI-BRATS 2015 challenge data are well registered, no alignment required. However, for practical clinic study, different MRI scans of the same patient often have mismatches at different time and/or with different scan settings. Thus, rigid and deformable registration [46, 47] will be needed. Secondly, for MRI images, there is no correction for geometric artifacts. In actual clinical study, the majority of the MRI images showing tissue regions are most likely affected by geometric artifacts. Wang et al. showed excellent data about susceptibility-induced geometric distortions [48], which is not negligible for the future tumor segmentation study. Therefore, without correcting geometric distortion within skull is likely to affect the robustness of all techniques in actual clinical data. This will also be part of our subsequent work to add some effective means to correct the geometric artifacts, such as a field-map based correction [49]. There is also the third practical task to study. The proposed technique is currently merely implemented by normal 2D and 3D U-Net architectures since the goal is to prove that the new ROI-aided strategy can improve the brain tumor segmentation accuracy. In our future work, we will test this ROI-aided strategy on other more advanced architectures such as the deeply supervised networks [50]. The next subsequent area of study is to develop a method that has the ability of automatic ROI searching with flexible ROI size. In the current study, the ROI size is fixed since the maximum tumor region is known in the data base. A practically valid strategy may need to have the ability to automatically identifying ROI sizes. In fact, the recent mask R-CNN has shown a great potential on flexible bounding box size for object detection [51], which may a direction in our future studies. Last but not the least, we have not yet applied multi-label segmentation for the MICCAI-BRATS 2015 challenge data. Multi-label segmentation shall also be part of our subsequent studies.

5 Conclusions

In summary, we develop a more accurate brain multi-modal tumor segmentation method using the new ROI-aided cascading network strategy. Numerical tests demonstrate that this technique can be a useful tool for brain cancer detection, diagnosis, and radiotherapy treatment planning.

Acknowledgements The authors would like to express their sincere gratitude to the anonymous referees and the editor for many valuable suggestions and comments that helped to improve the article. This work is partially supported by the National Science Foundation of USA (DMS-1615288).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Jemal A, Ward Elizabeth M, Johnson Christopher J et al (2017) Annual report to the nation on the status of cancer, 1975–2007,

Featuring tumors of the brain and other nervous system. J Natl Cancer Inst 103(9):714–736
2. Ray S, Bonafede MM, Mohile NA (2014) Treatment patterns, survival, and healthcare costs of patients with malignant gliomas in a large us commercially insured population. Am Health Drug Benefits 7(3):140–149
3. Ostrom Quinn T, Gittleman H, Truitt G et al (2018) CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015. Neuro-Oncol 20(4):1–86
4. Urbanska K, Sokolowska J, Szmidt M et al (2014) Glioblastoma multiforme—an overview. Contemporary Oncol-Termedia 18(5):307–312
5. Menze BH, Jakab A, Bauer S et al (2015) The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 34(10):1993–2024
6. Angulakshmi M, Lakshmi Priya GG (2017) Automated brain tumour segmentation techniques–a review. Int J Imaging Syst Technol 27(1):66–77
7. Işın A, Direkoğlu C, Şah M, (2016) Review of MRI-based brain tumor image segmentation using deep learning methods. Procedia Comput Sci 102:317–324
8. Saman S, Jamjala Narayanan S (2018) Survey on brain tumor segmentation and feature extraction of MR images. Int J Multimed Inf Retriev 8:79–99
9. Gordillo N, Montseny E, Sobrevilla P (2013) State of the art survey on MRI brain tumor segmentation. Magn Reson Imaging 31(8):1426–1438
10. Murthy TSD, Sadashivappa G (2014) Brain tumor segmentation using thresholding, morphological operations and extraction of features of tumor. In: 2014 International Conference on Advances in Electronics Computers and Communications, pp 1–6. https://doi.org/10.1109/ICAECC.2014.7002427
11. Ilhan U, Ilhan A (2017) Brain tumor segmentation based on a new threshold approach. Procedia Comput Sci 120:580–587
12. Zotin A, Simonov K, Kurako M et al (2018) Edge detection in MRI brain tumor images based on fuzzy C-means clustering. Procedia Comput Sci 126:1261–1270
13. Aslam A, Khan E, Beg MMS (2015) Improved edge detection algorithm for brain tumor segmentation. Procedia Comput Sci 58:430–437
14. Węgliński T, Fabijańska A (2011) Brain tumor segmentation from MRI data sets using region growing approach, Perspective Technologies and Methods in MEMS Design, pp 185–188
15. Charutha S, Jayashree MJ (2014) An efficient brain tumor detection by integrating modified texture based region growing and cellular automata edge detection. In: 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp 1193–1199. https://doi.org/10.1109/ICCICCT.2014.6993142
16. Shanthakumar P, Kumar PG (2015) Computer aided brain tumor detection system using watershed segmentation techniques. Int J Imaging Syst Technol 25(4):297–301
17. Khan MA, Lali IU, Rehman A et al (2019) Brain tumor detection and classification: a framework of marker-based watershed algorithm and multilevel priority features selection. Microsc Res Tech 82(6):909–922
18. Bauer S, Nolte L, Reyes M (2011) Segmentation of brain tumor images based on atlas-registration combined with a Markov-Random-Field lesion growth model. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp 2018–2021. https://doi.org/10.1109/ISBI.2011.5872808
19. Arakeri MP, Reddy GRM (2011) Efficient fuzzy clustering based approach to brain tumor segmentation on MR images. Comput Intell Inf Technol 250:790–795
20. Vijay J, Subhashini J (2013) An efficient brain tumor detection methodology using K-means clustering algoriftnn. In: 2013 International Conference on Communication and Signal Processing, pp 653–657. https://doi.org/10.1109/iccsp.2013.6577136
21. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. Br J Educ Psychol 24(6):417–520
22. Yang L, Xu Z (2019) Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning. Int J Mach Learn Cybernet 10:591–601
23. Kapas Z, Lefkovits L, Szilagyi L (2016) Automatic detection and segmentation of brain tumor using random forest approach. Model Decis Artif Intell 9880:301–312
24. Abdulbaqi HS, Mohd Zubir M, Omar AF et al (2014) Detecting brain tumor in Magnetic Resonance Images using Hidden Markov Random Fields and Threshold techniques. In: 2014 IEEE Student Conference on Research and Development, pp 1–5. https://doi.org/10.1109/SCORED.2014.7072963
25. Kumar TS, Rashmi K, Ramadoss S et al (2017) Brain tumor detection using SVM classifier. In: 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), pp 318–323. https://doi.org/10.1109/SSPS.2017.8071613
26. Luo Y, Yang B, Xu L et al (2018) Segmentation of the left ventricle in cardiac MRI using a hierarchical extreme learning machine model. Int J Mach Learn Cybernet 9:1741–1751
27. Liu X, Zhu T, Zhai L et al (2019) Mass classification of benign and malignant with a new twin support vector machine joint $l_{2,1}$ -norm. Int J Mach Learn Cybernet 10:155–171
28. Milletari F, Navab N, Ahmadi SA (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp 565–571. https://doi.org/10.1109/3DV.2016.79
29. Zhang J, Shen X, Zhuo T et al (2017) Brain tumor segmentation based on refined fully convolutional neural networks with a hierarchical dice loss. arXiv:1712.09093
30. Ben Naceur M, Saouli R, Akil M et al (2018) Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. Comput Method Program Biomed 166:39–49
31. Chen L, Wu Y, DSouza Adora M et al (2018) MRI tumor segmentation with densely connected 3D CNN. Medical Imaging 2018: Image Processing. vol. 10574. International Society for Optics and Photonics
32. Liu C, Gardner MS, Stephen J, Wen N et al (2019) Automatic segmentation of the prostate on CT images using deep neural networks (DNN). Int J Radiat Oncol Biol Phys 104(4):924–932
33. Yan K, Wang X, Kim J et al (2019) A propagation-DNN: deep combination learning of multi-level features for MR prostate segmentation. Comput Methods Programs Biomed 170:11–21
34. Ito R, Nakae K, Hata J et al (2019) Semi-supervised deep learning of brain tissue segmentation. Neural Netw 116:25–34
35. Feng X, Qing K, Tustison NJ et al (2019) Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. Med Phys 46(4):2169–2180
36. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. Int Conf Med Image Comput Comput-Assist Interv 9351:234–241
37. Zhou C, Ding C, Wang X et al (2020) One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. IEEE Trans Image Process 29:4516–4529
38. Kayalibay B, Jensen G, van der Smagt P (2017) CNN-based segmentation of medical imaging data. arXiv:1701.03056
39. Isensee F, Kickingereder P, Wick W et al (2018) Brain tumor segmentation and radiomics survival prediction: contribution

to the brats 2017 challenge. International MICCAI Brainlesion Workshop. Springer, Cham, pp 287–297
40. Kamnitsas K, Ledig C, Newcombe VFJ et al (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 36:61–78
41. Qin Y, Kamnitsas K, Ancha S et al (2018) Autofocus layer for semantic segmentation. International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 603–611
42. Ranjbarzadeh R, Ghoushchi SJ, Bendechache M et al (2021) Lung Infection Segmentation for COVID-19 Pneumonia Based on a Cascade Convolutional Network from CT Images. BioMed Res Int 5544742:16. https://doi.org/10.1155/2021/5544742
43. Zhao W, Vernekohl D, Zhu J et al (2016) A model-based scatter artifacts correction for cone beam CT. Med Phys 43(4):1736–1753
44. Altunbas C, Lai CJ, Zhong Y et al (2014) Reduction of ring artifacts in CBCT: Detection and correction of pixel gain variations in flat panel detectors. Medical Phys 41(9):091913
45. Gai S, Zhang B, Yang C et al (2018) Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution. Digit Signal Process 72:192–207
46. Lin JS, Fuentes DT, Chandler A et al (2017) Performance Assessment for Brain MR Imaging Registration Methods. Am J Neuroradiol 38(5):973–980
47. Padgett KR, Stoyanova R, Pirozzi S et al (2018) Validation of a deformable MRI to CT registration algorithm employing same day planning MRI for surrogate analysis. J Appl Clin Med Phys 19(2):258–264
48. Wang H, Balter J, Cao Y (2013) Patient-induced susceptibility effect on geometric distortion of clinical brain MRI for radiation treatment planning on a 3T scanner. Phys Med Biol 58(3):465–477
49. Togo H, Rokicki J, Yoshinaga K et al (2017) Effects of field-map distortion correction on resting state functional connectivity MRI. Front Neurosci 11:656–656
50. Dou Q, Lequan Yu, Chen H et al (2017) 3D deeply supervised network for automated segmentation of volumetric medical images. Med Image Anal 41:40–54
51. He K, Gkioxari G, Dollár P et al (2017) Mask R-CNN, 2017 IEEE international conference on computer vision (ICCV), pp 2980–2988