Asymmetric Ensemble of Asymmetric U-Net Models for Brain Tumor Segmentation With Uncertainty Estimation

Sarahi Rosas González

UMR Inserm U1253, iBrain, Université de Tours, Inserm, Tours, France

Abstract

Accurate brain tumor segmentation is crucial for clinical assessment, follow-up, and subsequent treatment of gliomas. While convolutional neural networks (CNN) have become state of the art in this task, most proposed models either use 2D architectures ignoring 3D contextual information or 3D models requiring large memory capacity and extensive learning databases. In this study, an ensemble of two kinds of U-Net-like models based on both 3D and 2.5D convolutions is proposed to segment multimodal magnetic resonance images (MRI). The 3D model uses concatenated data in a modified U-Net architecture. In contrast, the 2.5D model is based on a multi-input strategy to extract low-level features from each modality independently and on a new 2.5D Multi-View Inception block that aims to merge features from different views of a 3D image aggregating multi-scale features. The Asymmetric Ensemble of Asymmetric U-Net (AE AU-Net) based on both is designed to find a balance between increasing multi-scale and 3D contextual information extraction and keeping memory consumption low. Experiments on 2019 dataset show that our model improves enhancing tumor sub-region segmentation. Overall, performance is comparable with state-of-the-art results, although with less learning data or memory requirements. In addition, we provide voxel-wise and structure-wise uncertainties of the segmentation results, and we have established qualitative and quantitative relationships between uncertainty and prediction errors. Dice similarity coefficient for the whole tumor, tumor core, and tumor enhancing regions on BraTS 2019 validation dataset were 0.902, 0.815, and 0.773. We also applied our method in BraTS 2018 with corresponding Dice score values of 0.908, 0.838, and 0.800. Glioma is the most frequent primary brain tumor (1). It has its origin in glial cells and can be classified into I to IV grades, depending on phenotypic cell characteristics. In this grading system, low-grade gliomas (LGGs) correspond to grades I and II, whereas high-grade gliomas (HGGs) are grades III and IV. The primary treatment is surgical resection followed by radiation therapy and/or chemotherapy.MRI is a non-invasive imaging technique commonly used for diagnosis, surgery planning, and follow-up of brain tumors due to its high resolution on brain structures. Currently, tumor regions are segmented manually from MRI images by radiologists, but due to the high variability in image appearance, the process is very time consuming and challenging, and inter-observer reproducibility is considerably low (2). Since accurate tumor segmentation is determinant for surgery, follow-up, and subsequent treatment of glioma, finding an automatic and reproducible solution may save time for physicians and contribute to improving the clinical assessment of glioma patients. Based on this observation, the Multimodal Brain Tumor Segmentation Challenge (BraTS) aims at stimulating the development and the comparison of the state-of-the-art segmentation algorithms by making available an extensive pre-operative multimodal MRI dataset provided with ground truth labels for three tumor tissues: enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core. This dataset contains four modalities: T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR), T1-weighted (T1), and T1 with contrast-enhancing gadolinium (T1c) (3–7).Modern convolutional neural networks (CNNs) are currently state-of-the-art in many medical image analysis applications, including brain tumor segmentation (8). CNNs are hierarchical groups within filter banks that extract increasingly high-level image features by feeding the output of each layer to the next one. Recently, Ronneberger et al. (9) proposed an effective U-Net model, a fully convolutional network (FCN) encoder/decoder architecture. The encoder module consists of multiple connected convolution layers that aim to gradually reduce the spatial dimension of feature maps and capture high-level semantic features appropriate for class discrimination. The decoder module uses upsampling layers to recover the spatial extent and object representation. The main contribution of U-Net is that, while upsampling and going deeper into the network, the model concatenates the higher resolution features from the encoder path with the upsampled features in the asymmetric decoder path to better localize and learn representations in following convolutions. The U-Net architecture is one of the most widely used for brain tumor segmentation, and its versatile and straightforward architecture has been successfully used in numerous segmentation tasks (10–14). All top-performing participants in the last two editions of the BraTS challenge used this architecture (15–22).While 3D

CNN can provide global context information of volumetric tumors, the large size of the images makes the use of 3D convolutions very memory demanding, which limits the patches and batch size, as well as the number of layers and filters that can be used (23). Consequently, the use of 2D convolutions for slice-by-slice segmentation is also a common practice that reduces memory requirement (24). Multi-view approaches have also been developed to address the same problem. McKinley et al. (18) and Xue et al. (25) showed that applying 2D networks in axial, sagittal, and coronal views and combining their results can recover 3D spatial information. Recently, one of the top-performing submissions in the BraTS 2019 challenge (20) proposed a hybrid model that goes from 3D to 2D convolutions extracting two-dimensional features in each of the orthogonal planes and then combines the results in an ensemble model. Wang et al. (16) demonstrated that using three 2.5D networks to obtain separate predictions from three orthogonal views and fusing them at test time can provide more accurate segmentations than using an equivalent 3D isotropic network. While they require the training and optimization of several models, ensemble models are currently the top-performing methods for brain tumor segmentation.On the other hand, some strategies have been implemented to aggregate multi-scale features. Cahall et al. (26) showed a significant improvement in brain tumor segmentation by incorporating Inception blocks (27) into a 2D U-Net architecture. Wang et al. (16) and McKinley et al. (20) used dilated convolutions (28) in their architecture with the same aim of obtaining both local and more global features. While significant, aggregating multi-scale features is limited by the requirement of more memory capacity. To address this, the use of Inception modules has been incorporated into 2D networks (26), not taking advantage of 3D contextual information. In addition, Inception modules have been integrated into a cascade network approach (29). The model first learns the whole tumor, then the tumor core, and finally the enhancing tumor region. This method requires three different networks and thus increases the training and inference time. Another approach to extract multi-scale features uses dilated convolutions. This operation was explicitly designed for semantic segmentation and tackled the dilemma of obtaining multi-scale aggregation without losing full resolution, increasing the receptive field (28). Wang et al. (16) and McKinley et al. (20) implemented different dilation rates in sequential convolutions; nevertheless, it has not been used to extract multi-scale features in a parallel way in a single layer and, if not applied carefully, can cause gridding effects, especially in small regions (30).In terms of accuracy and precision, the performance of CNNs are currently comparable with human-level performance or even better in many medical image analysis applications (31). However, CNNs have also often been shown to produce inaccurate and unreliable probability estimates (32, 33). This has drawn attention to the importance of uncertainty estimation in CNN (34). Among other advantages, the measurement of uncertainties would enable knowing how confident a method is in implementing a particular task. This information can facilitate CNN's incorporation into clinical practice and serve the end-user by focusing attention on areas with high uncertainty (35).In this study, we propose an approach that addresses a current challenge of brain tumor segmentation, keeping reduced memory requirements while benefiting from multi-scale 3D information. To do so, we propose an ensemble model, called Asymmetric Ensemble Asymmetric U-Net (AE AU-Net), based on an Asymmetrical 3D residual U-Net (AU-Net) using two different kinds of inputs: (1) concatenated multimodal 3D MRI data (3D AU-Net) and (2) a 2.5D Multi-View Inception Multi-Input module (2.5D AU-Net). The proposed AU-Net is wider in the encoding path to extract more semantic features, has residual blocks in each level to increase training speed, and additive skip connections between the encoding and decoding path instead of a concatenation operation to reduce the memory consumption. The proposed 2.5D strategy allows us to extract low-level features from each modality independently. In this way, the model can retrieve specific details related to tumor appearance from the most informative modalities, without the risk of being lost when combined during the downsampling. The Multi-View Inception block aims to merge features from both different views and different scales simultaneously, seeking a balance between 3D information usage and memory footprint. In addition, we use an ensemble of models to improve our segmentation results and the generalization power of our method and also as a way to measure epistemic uncertainty and estimate structure-wise uncertainty.