Multimedia Tools and Applications (2021) 80:27069–27094
https://doi.org/10.1007/s11042-021-10915-y

RD^{2}A: densely connected residual networks using ASPP for brain tumor segmentation

Parvez Ahmad[1] · Hai Jin[1] · Saqib Qamar[1] · Ran Zheng[1] · Adnan Saeed[2]

Received: 10 July 2020 / Revised: 28 October 2020 / Accepted: 1 April 2021 /
Published online: 13 May 2021

Abstract

The variations among shapes, sizes, and locations of tumors are obstacles for accurate automatic segmentation. U-Net is a simplified approach for automatic segmentation. Generally, the convolutional or the dilated convolutional layers are used for brain tumor segmentation. However, existing segmentation methods of the significant dilation rates degrade the final accuracy. Moreover, tuning parameters and imbalance ratio between the different tumor classes are the issues for segmentation. The proposed model, known as *Residual-Dilated Dense Atrous-Spatial Pyramid Pooling* (RD^{2}A) 3*D* U-Net, is found adequate to solve these issues. The RD^{2}A is the combination of the residual connections, dilation, and dense ASPP to preserve more contextual information of small sizes of tumors at each level encoder path. The multi-scale contextual information minimizes the ambiguities among the tissues of the *white matter* (WM) and *gray matter* (GM) of the infant's brain MRI. The BRATS 2018, BRATS 2019, and iSeg-2019 datasets are used on different evaluation metrics to validate the RD^{2}A. In the BRATS 2018 validation dataset, the proposed model achieves the average dice scores of 90.88, 84.46, and 78.18 for the *whole tumor*, the *tumor core*, and the *enhancing tumor*, respectively. We also evaluated on iSeg-2019 testing set, where the proposed approach achieves the average dice scores of 79.804, 77.925, and 80.569 for the *cerebrospinal fluid* (CSF), the *gray matter* (GM), and the *white matter* (WM), respectively. Furthermore, the presented work also obtains the mean dice scores of 90.35, 82.34, and 71.93 for the *whole tumor*, the *tumor core*, and the *enhancing tumor*, respectively on the BRATS 2019 validation dataset. Experimentally, it is found that the proposed approach is ideal for exploiting the full contextual information of the 3*D* brain MRI datasets.

Keywords Deep learning · Residual network · Dense connections · Brain tumors · Atrous-spatial pyramid pooling

✉ Hai Jin
hjin@hust.edu.cn

[1] National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China

[2] School of Hydropower and Information Technology, Huazhong University of Science and Technology, Wuhan, China

1 Introduction

Gliomas can affect the normal working of the human brain. Gliomas can be categorized into two grades: *high-grade glioblastoma* (HGG) and *low-grade glioblastoma* (LGG) [20], in which each grade has the number of classes. Based on the grading system, it is crucial to make the prediction and rate of the tumors. The MRI technique is mainly used for an in-depth analysis of the brain structure. However, the unusual variations among the shapes, size, and location of a tumor [23] in the MRIs are obstacles to developing an efficient, accurate algorithm. Feature learning has the great potential to handle such kind of problem. Classical existing methods also delivered excellent results. However, in the traditional approach, the unfeasible study of brain tumors makes impractical. In comparison to them, the process of feature learning gives an abstract representation of data. Figure 1 shows the process of feature representation with the Alexnet architecture [28]. The modified Alexnet architecture has five convolution layers to learn features. A pooling layer is employed between every pair of convolution layers to reduce the original input resolution. Furthermore, two convolutional layers of kernel size of 1×1 are used to reduce the features. The reduced features of the last convolution layer match the number of labels of a dataset. Finally, the output reflects the probability distribution of different classes or labels by using the softmax.

Figure 1 shows a 2D architecture in which each convolution layer, containing several 2D filters or kernels. Each filter yields the feature maps when applied to the channels of the previous layer. For example, 24 filters in the first convolution layer after convolving the input of 3 channels yield output feature maps $24 \times 128 \times 128$. These output feature maps then input to the next convolution layer and so on. To drive the generalized formula for the output feature maps of each layer l, let C_l denoting the convolution filters or kernels and p_{l-1}^{z} denoting the 2D array corresponding to the z^{th} input. Then the output feature map of each kernel of layer l is

$$q_l^{kernel} = f\left(\sum_{z=1}^{C_{l-1}} Weight_l^{kernel,z} \star p_{l-1}^{z} + b_l^{kernel}\right) \tag{1}$$

where $Weight_l^{kernel,z}$ denoting the weight of each kernel, b_l^{kernel} denoting a bias, and $\star$ representing convolution operation. f is a Leaky Rectified Linear Unit (Leaky ReLU) nonlinear activation function. It has α as an extra parameter to prevent the problem of zero gradients during the training. Mathematically, it can be written as

$$f\left(x_j\right) = x_j + \alpha x_j = \begin{cases} x_j & \text{if } x_j > 0 \\ \alpha x_j & \text{if } x_j \leq 0 \end{cases} \tag{2}$$

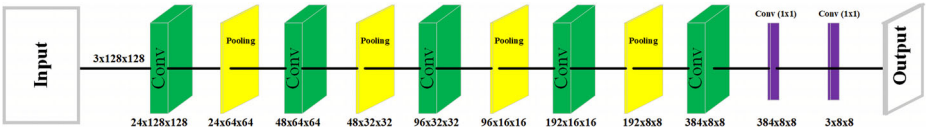

Fig. 1 Feature representation of the modified Alexnet architecture. Each convolution layer (green) is of a kernel size of 3×3. A stride of size two is used with each pooling layer (yellow) to reduces the input resolution. The last two convolution layers reduce the number of channels or features before the final output

where x_j denoting the input features and $f\left(x_j\right)$ representing the output features. The spatial dimensions of the output feature maps of each convolution layer reduce by a pooling layer. This spatial reduction is possible by using a stride of size 2 in the pooling layer. Mathematical expression for each pooling layer can be

$$y_l^{kernel} = \max\left(q_l^{kernel}\right) \tag{3}$$

where y_l^{kernel} denoting feature maps of pooling layer l for the $kernel^{th}$ input feature map $\left(q_l^{kernel}\right)$, max (.) denoting the max-pool operation. Finally, the softmax activation is performed on the reduced feature maps of the last 1×1 convolution to generate the probability distribution of different classes. Mathematically, the softmax activation is

$$O_c = \frac{\exp\left(q_L^c\right)}{\sum_{c'=1}^{C} \exp\left(q_L^{c'}\right)} \tag{4}$$

where the classes are denoting by the C and the last layer is representing by L.

Figure 1 reflects the deep Convolutional Neural Networks (CNNs). However layers in the deep models have complex inter-connections for better learning. A Fully Convolutional Neural Network (FCNN), especially, U-Net is very popular in biomedical image segmentation [42, 55]. The U-Net architecture gains popularity due to skip-connections [39, 56]. Skip-connections perform the concatenation operation on the maps of the different parts of a U-Net model. The potential of skip-connections can be understood from basic FCN architecture [30] (Fig. 2a) and different variations of U-Net models (Fig. 2b, and c). In Fig. 1, the feature maps of each convolution layer are reduced by the max-pooling layer. Therefore, the size of the output is smaller than that of the input. This problem can be resolved by applying the upsampling layer. For example, on upsampling the last pooling layer (in Fig. 2a) at 32, the resulting output size is equal to the input patch. Furthermore, the part with the pooling layers is known as the encoder, while the upsampling part is the decoder. However, limited contextual information is a critical issue in the deeper layers, which can be addressed by combining the predictions of the different layers (see $FCN-8s$ in Fig. 2a). In this way, the skip-connections combine in-depth, coarse, semantic information of the decoder part with the encoder part's shallow, adequate location information. The FCN architecture is depicted in Fig. 2a. All the U-Net architectures, either with non-residual convolution blocks (Fig. 2b) or with residual convolution blocks (Fig. 2c), are followed the basic design of the FCN. However, the lack of non-linearities in the decoder part ruled out the use of FCN for the medical image segmentation. In the meantime, U-Net is successfully resolving the deficiency of the non-linearities by adding the convolution between the upsampling layers (see lower part of Fig. 2b and c). Furthermore, U-Net models' strength is improved with the concatenation operation instead of the simple addition operation (see $FCN-8s$ in Fig. 2a). Simultaneously, depth is an essential criterion for U-Net models (for example, Fig. 2b) to improve segmentation accuracy. However, the gradients may be vanishing with such deep models. The vanishing gradients' problem is resolved by adding the residual blocks in the U-Net (Fig. 2c). Furthermore, a deep study of U-Net architectures for medical image segmentation can be accessed by a recently published survey paper [29]. The mathematical notations of convolutions and upsampling layers are defined in Section 3.

Traditional U-Net models either with the non-residual (see Fig. 2b) or residual (see Fig. 2c) or dense blocks [22] perform a sequence of convolution and strided convolution or max-pool operations on the original images, which can reduce the spatial dimensions of input images and increase the receptive field size in the sub-sampling process. While

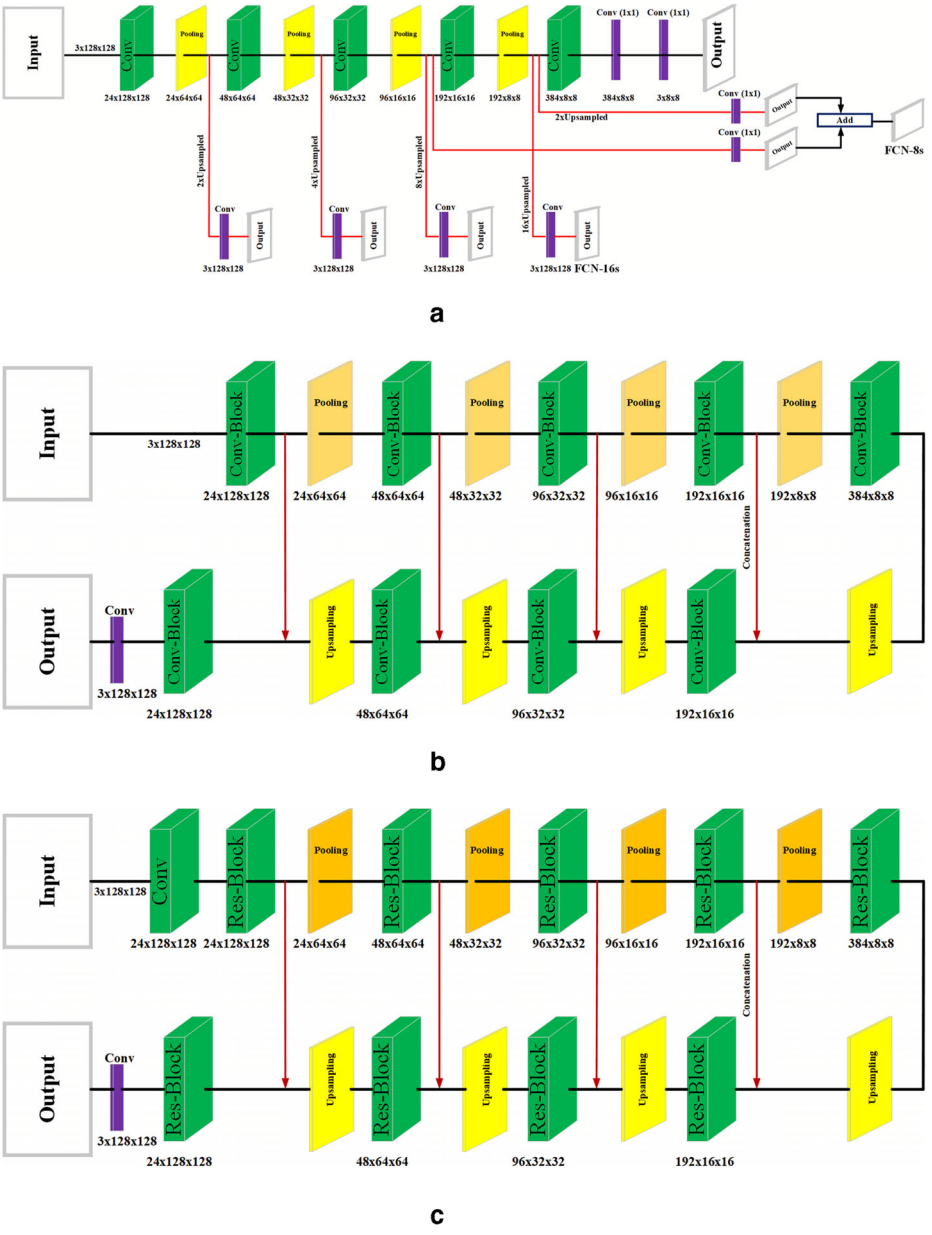

Fig. 2 Outline of FCN and U-Net models for medical image segmentation. **a** FCN **b** U-Net without residual block **c** U-Net with residual block

the upsampling operation recovers the reduced size of the input images; however, the possibility of losing the critical information of original images by the sub-sampling process can not be completely ruled out. To prevent this loss, dilated convolutions [53] can be used to learn more contextual information on the encoder path. The similarity between dilated and ordinary convolutions is that the convolution core's size is the same. At the same time,

the dilated convolution increases the receptive field size; however, the number of training parameters does not increase. Therefore, the dilated based models are not only limited to the natural images, but they are also continuously improving the segmentation accuracy in the medical domain [12, 13]. Furthermore, Devalla et al. [11] presented a U-Net model with equal or larger rates of dilation in the different parts to learn more contextual information. However, large or equal dilation rates introduces the gridding problem [54] (Fig. 3a). A similar problem of large dilation rates also exists with *atrous spatial pyramid pooling* (ASPP) [9]. ASPP in which multiple dilated layers have a parallel arrangement, information of these multiple-scales, also known as multi-scale information, can further boost the segmentation accuracy. Dilation rate increases the receptive field size by inserting the extra zeros between the kernel elements. This gap is continuously widening when a series of convolution layers have either similar or larger dilation rates. Here, the sparsed kernel of convolutions fails to capture any local information. In this way, the result of the gridding problem is the local information's complete loss. This loss of local information may degrade the final accuracy. However, the gridding effect can be minimized using different dilation rates [48] (Fig. 3b). Moreover, in U-Net architecture, the mapping of information between encoder and decoder

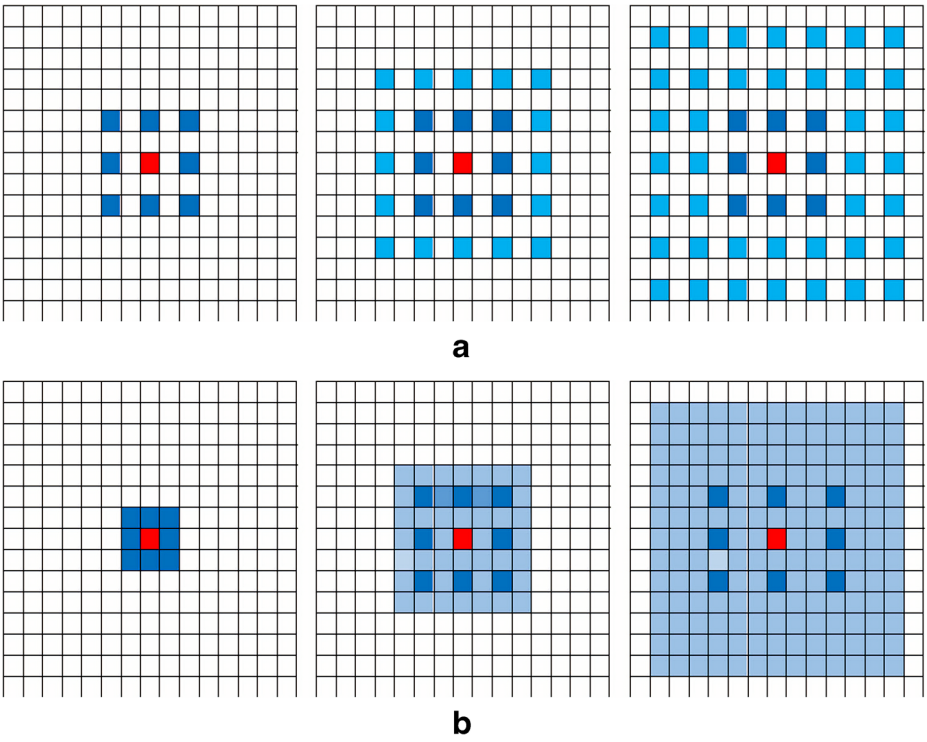

Fig. 3 An instance of gridding problems. **a** depicts three convolution layers with equal dilation rates (r = 2) with kernel size 3×3. The pixels (receptive field, denoting by blue) contribute to calculating the center pixel (representing by red). The similar dilation rates through three convolution layers reduce the local information because zeros are inserting between pixels. This effect is diminished when **b** different dilation rates (r = 1, 2, 3) are used. These rates are preventing the checkerboard patterns, which are introducing due to equal dilation in (see a) or large dilation rates in ASPP [9]. Source [48]

parts is different, which arises the semantic gap in the architecture. It can reduce the functionality of U-Net models. Zhou et al. [56] proposed to minimize the semantic gap by improving skip-connections. However, this improvement increases the complexity due to the multiple paths between the encoder-decoder sub-networks.

In the above discussion, U-Net architectures have followed 2D convolutions, while, several researchers extended their works into 3D convolutions [1, 20, 46] with excellent results to solve the problem of the brain MRI segmentation. In this work, we also add depth to a $3D$ U-Net model by employing the residual connections [15] to resolve the issue of vanishing gradients. However, residual connections used element-wise addition operations that gives a limited improvement in the segmentation accuracy. At the same time, concatenation operations are known to improve the width of the channels. Therefore, we also employed concatenations to fuse the features of different sub-networks of a 3D U-Net and in *atrous-spatial pyramid pooling* (ASPP) [9] blocks. However, to concatenate the features of multiple scales in ASPP, we used the recently proposed dense connections [19]. While dense connections resolve the vanishing gradients problem, they also offered features reusability property by concatenating all the layers' feature maps. That means the input to a particular layer simultaneously has the coarse, semantic information of the deeper layers, and shallow, adequate location information of the lower layers. This information is further improved in the deeper layers. Furthermore, using an appropriate growth-rate, dense connections reduce the parameters generated by the residual networks. This reduction is essential to demonstrate maximum accuracy with minimum learned parameters. Moreover, we adopt 3D dilated convolutions and preventing the gridding problem by employing different rates of dilations. Finally, we use dense ASPP blocks on the skip-connections' output feature maps to learn multi-scale features to improve the segmentation outcomes. This multi-scale learning from the redesigned skip-connections also minimizes the semantic gap without introducing complexity.

Inspired by the success of the residual and dense connections, dilation, and the ASPP techniques, we have proposed a variant form of $3D$ U-Net with the combination of the residual connections, dilation, and dense ASPP. We have offered an RD^2A (Residual-Dilated Dense ASPP) $3D$ U-Net model. The key contributions of this study are given below:

- A variant form of $3D$ U-Net. We used combined approach of residual connections and densely connected ASPP.
- To avoid possible loss of information during training in the proposed model, we choose appropriate rates of dilation layer to gain the proper size of the receptive field on BRATS datasets. Additionally, we used dense connections among the multiple sizes of the receptive field in ASPP on the feature maps of a residual-dilated 3D U-Net model for exploiting the full contextual information of the 3D brain MRIs datasets.
- We have worked on BRATS 2018 and BRATS 2019 datasets, where the proposed model achieved state-of-the-art performances compared to other recent methods in terms of both parameters and accuracy.
- We have worked on iSeg-2019 datasets and achieved the best scores on the testing dataset against the best method of the iSeg-2019 validation dataset.

2 Related work

U-Net [39] introduces the concept of skip-connections. Such connections are useful to preserve the original information at each level of the encoder. At the decoder, the information

concatenates with its predecessor's level information. These connections open the door of a deep network to better understand biomedical images' complex structure by using local and global contextual information. Marcinkiewicz et al. [33] used cascaded U-Net for brain segmentation, in which they used the first step works as a detection while the second as a multi-class classifier. Hu et al. [17] proposed a fusion method to concatenate the features of three 2D U-Net networks. Chen et al. [10] replaced the block of a residual 3D U-Net with the inception block. Two layers replace each 3D layer in a block: one for spatial information and the other for channel representation [50]. Chen et al. [10] presented their work with the parameters reduction of a 3D convolution. However, huge parameters exist due to the increased number of layers. Kermi et al. [26] proposed a 2D FCN to resolve the high memory demands of 3D brain MRI. Residual connections are used to build a profound network; therefore, the resulting model generates many parameters during training. A combination of two cost functions was used to balance the different classes. Isensee et al. [20] proposed a residual encoder-decoder architecture with a 3D UpSampling layer on the resulting maps of different blocks at decoder sub-network to extract the deeper features. Xu et al. [51] proposed a segmentation task with the combination of three 3D encoder-decoder models. The output of the first model worked like the input of the next architecture. Wang et al. [46] proposed similar approach to Xu et al. [51] without end-to-end networks. Mehta et al. [34] proposed a 3D U-Net similar to Isensee et al. [20], they used transposed convolution instead of the UpSampling layer. Roy et al. [40] implemented large dilation rates in ASPP. However, the complexity was a major problem with the combined orthogonal networks. Ensembling of different models was also proposed to improve the segmentation results. Kori et al. [27] and Kamnitsas et al. [23] implemented the idea of ensembling on different models with a majority voting scheme.

As mentioned earlier, the profound variations of U-Net architectures can learn significant features about unhealthy brain structures. Hence, the most straightforward strategy with deep U-Net architectures is residual learning. The residual networks have several direct connections between layers to prevent the problem of vanishing gradients. Therefore, nearly all the above-discussed methods used residual connections. However, the generation of huge parameters is a severe problem with residual networks. Furthermore, designing skip-connections with the traditional approach hinders U-Net architectures' potential from learning sophisticated information. Moreover, the system's complexity increases with the combination of various architectures, such as cascaded U-Nets, ensembling, etc. In our work, we also used the residual network for the encoder sub-part. Here, dilated convolutional layers are employed. In this way, our redesigned encoder sub-network can learn more contextual information from input brain MRIs than the traditional U-Net architectures. Furthermore, we used dense ASPP blocks to design the skip-connections to allow our network to learn more fine-grained multi-scale features. The dense connections reduced the parameters and offered more scaling [52] to each dilated convolution by improving the channels' width. This scaling factor in multi-scale features minimizes the semantic gap. In this way, our proposed architecture without adding any complexity can solve all the previously proposed methods' problems. Furthermore, a comparison between the proposed dense ASPP blocks and the architecture of Zhou et al. [56] is depicted in Fig. 4.

3 RD^{2}A (residual-dilated dense ASPP 3D U-Net)

Figure 5 shows our proposed architecture. The combined approach of the residual connections, dilation, and dense ASPP consists of a Residual-Dilated and Dense ASPP blocks.

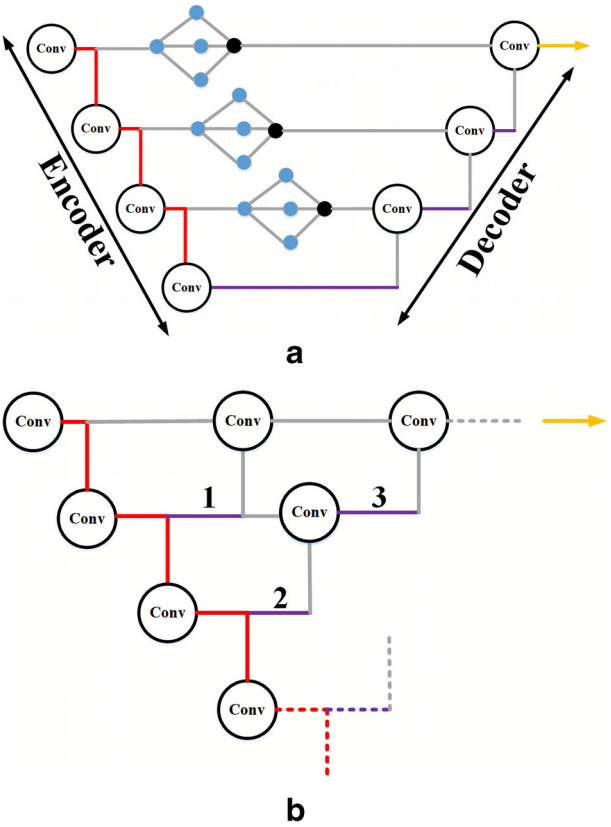

Fig. 4 Outline of skip-connections variations, **a** illustrates the proposed dense ASPP blocks on the skip-connections. Conv in an oval shape denoting the standard convolution layers. Simultaneously, the solid red and blue lines represent the pooling and the upsampling operations while their dashed forms denote the repetitive pooling and the upsampling processes in **b**. In the meantime, the concatenation operations are indicating in gray. Furthermore, small circular shapes (blue) representing the different dilation rates of the convolutions, and the black oval shapes denoting the $1 \times 1 \times 1$ convolutions. The concatenated operations of multiple dilated convolutions deduced more fine-grained multi-scale features from high and low-level input resolutions (indicating by the encoder). These multi-scale features further improved in the decoder part. **b** illustrates the UNet++ in detail. The numbers represent the concatenation operations between multiple encoder-decoder sub-networks, which will be continuously increasing by adding depth to the network. As a consequence, redesigned skip-connections in UNet++ increases complexity. Moreover, the solid orange line in **a** and **b** represent the output

Residual-Dilated blocks are in the first part of the RD^{2}A $3D$ U-Net model and the application of dense ASPP on the feature maps of the Residual-Dilated blocks. The residual-dilated block shares a common idea of a dilated convolutional layer. Figure 6 shows the design of the residual-dilation block. To learn more contextual information, we have used dilated convolutional with the rates 1 and 2 in each residual-dilation block. Unlike Wang et al. [46], we implement the different dilation rates within each residual-dilation block because the same rates introduce the gridding problem [48]. In the proposed architecture, we have used 4 residual-dilation blocks. After each block, strided convolutional is used to reduce the input resolution. The dense ASPP block is applied to the feature maps of the first three residual-dilation blocks before the sub-sampling layers. Therefore, the reduced size of feature maps

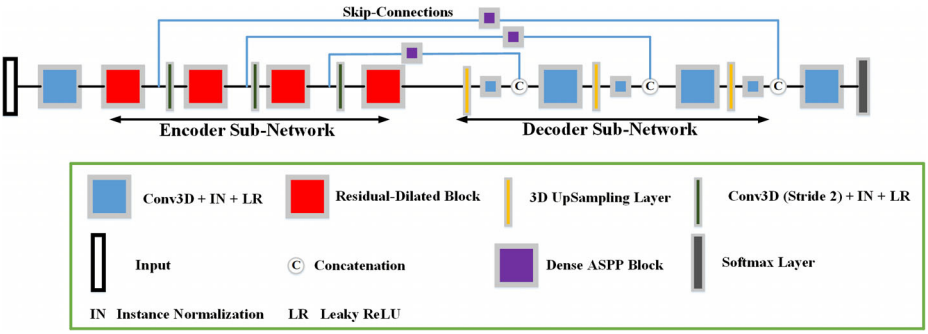

Fig. 5 Proposed architecture. RD^{2}A 3D U-Net model is divided into a lower part (encoder sub-network) and a higher part (decoder sub-network). The lower part has the residual-dilated blocks (red). For each residual-dilated block (Fig. 6), two different rates of dilated convolutional layers are used. The dense ASPP blocks (purple) are employed after each upsampling layer (violet) on the maps of the lower part in the higher part before each concatenation operation (symbol C in an oval shape). For each Dense ASPP block (Fig. 7), dense connections are used among the three different parallel dilated convolutional layers

is processed via the multiple parallels dilated layers on different rates with dense connections to exploit the multi-scale contextual information, and non-dilated layer implemented to deduce global contextual information. Figure 7 exhibits the dense ASPP block of our proposed architecture. Here, we used four different dilated layers with rates of 1, 2, 3, and 5. The residual-dilation and dense ASPP blocks exist at the encoder part of our proposed architecture. In the decoder path, the input resolution at the corresponding level's predecessor of encoder path recovers by using a 3D UpSampling layer of size $2 \times 2 \times 2$. In our proposed approach, each level of the encoder path preserves more contextual information with the concatenation of generated features at each level of the decoder path.

We divide the proposed architecture into two parts: residual-dilation blocks at encoder R_E and the upsampling process at decoder R_U. We use 3D convolutional filters or kernels to transform the 3D raw brain MRIs into our architecture features. For the first convolutional layer, when a kernel of size $3 \times 3 \times 3$ (denoting by $Weight_1^{1,4}$) is applied to the input patch of

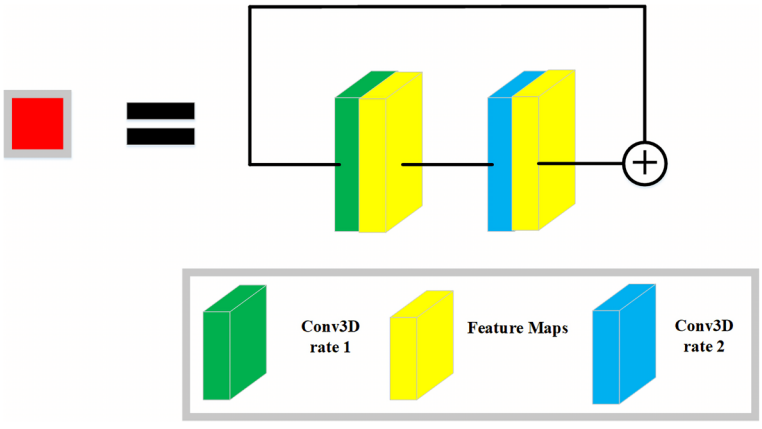

Fig. 6 Residual-dilation block. The two different dilation rates are used to enhance the sizes of the receptive field in each block

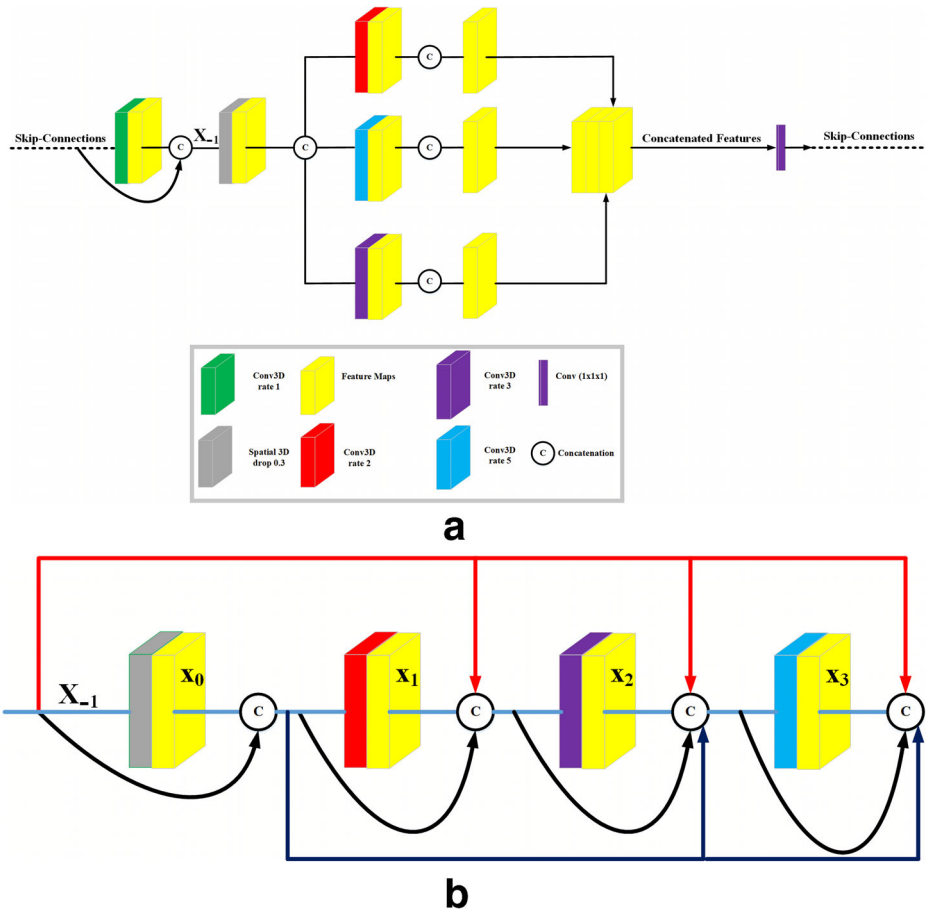

Fig. 7 Outline of a dense ASPP block, **a** illustrates a dense ASPP block in detail, **b** demonstrate the idea of the concatenation operations (denoting by a symbol c in oval shape) in the parallel dilated layers. The input feature maps to a layer is the combination of output feature maps of all previous layers. For example, the input to a last dilated convolution (blue) includes the output feature maps, x_{-1}, x_0, x_1 and x_2 because of the dense connections

size $128 \times 128 \times 128$ (representing by $[I_F]_{-1}$), then feature map of size $1 \times 128 \times 128 \times 128$ is extracted. 4 are the number of modalities or channels which are used to extract the input patches. Since 24 filters (denoting by $\sum_{z=1}^{4} Weight_1^{24,4}$) are in use. Hence, $24 \times 128 \times 128 \times 128$ (denoting by $[I_F]_0$) will be the final size of the features for the first convolutional layer. Mathematically, the procedure of features extraction for the first weighted layer can be written as

$$[I_F]_0 = f\left(\sum_{z=1}^{4} Weight_1^{kernel,z} \star [I_F]_{-1} + b_1^{kernel}\right) \tag{5}$$

where b_1^{kernel} is bias term. $\star$ representing convolution operation. f is a Leaky Rectified Linear Unit (Leaky ReLU) non-linear activation function which was defined in (2).

$[I_F]_0$ will be input to our first R_E to generate the feature maps $[R_E]_{F_0}$ in (6). Since R_E denoting a residual-dilated block of two convolutional layers. Hence, for each layer of R_E, we use a similar idea of feature extraction, which was explained in (5).

$$[R_E]_{F_0} = \begin{cases} f\left(\sum_{z=1}^{24} Weight_2^{kernel,z} \star [I_F]_0 + b_2^{kernel}\right) \\ f\left(\sum_{z=1}^{24} Weight_3^{kernel,z} \star [I_F]_1 + b_3^{kernel}\right) \end{cases} \tag{6}$$

where $Weight_2^{kernel,z}$ and $Weight_3^{kernel,z}$ are the filters of the second and third convolutional layers, respectively. $[I_F]_0$ and $[I_F]_1$ are the input features of sizes $24 \times 128 \times 128 \times 128$.

By taking inspiration from (6), the generalized formula of the feature maps for the remaining Rs_E of the encoder sub-part can be written as

$$[R_E]_{F_D} = f\left(\sum_{z=1}^{C_{l-1}} Weight_l^{kernel,z} \star [R_E]_{F_{(D-1)}} + b_l^{kernel}\right) \tag{7}$$

where $D = 1, 2, 3$ denoting the current R_E of the encoder. For each R_E, C_l representing number of filters of a layer l where l=1, 2, b_l^{kernel} representing the bias term of layer l, and $[R_E]_{F_D}$ denoting feature maps. In the meantime, $[R_E]_{F_{(D-1)}}$ representing the input maps of previous Rs_E.

After the feature extraction from the encoder part, we perform the up-sampling operation R_U in the decoder part. Here, the final input resolution of encoder path upsamples to its predecessor's input size. To explain the current R_U, consider a tensor of the shape $192 \times 16 \times 16 \times 16$ representing the previous convolutional layer's feature maps. These feature maps are resized on applying an UpSampling layer of factor $2 \times 2 \times 2$. After this step, the size of feature maps at current R_U is twice than that of its corresponding R_E, i.e., $192 \times 32 \times 32 \times 32$ at R_U and $96 \times 32 \times 32 \times 32$ at R_E. To exactly matches the sizes of features at both ends, a 3D convolution with filter size $3 \times 3 \times 3$ is applied on R_U. After then, a dense ASPP block of three parallel dilated layers is applied on R_E. Finally, two tensors (one of R_U and other of R_E) of the shapes $96 \times 32 \times 32 \times 32$ are concatenated. A 3D convolutional layer is then applied to the combined tensor of the shape $96 \times 32 \times 32 \times 32$ for the remaining Rs_U. The general formula for each level of R_U can be summarized in (8a)

$$[R]_{U_D} = UpSampling3D\,(2 \times 2 \times 2)\,([R_E]_{F_D}) \tag{8a}$$

$$[R]_{U_D} = f\left(\sum_{z=1}^{C_{l-l}} Weight_l^{kernel,z} \star [R_E]_{F_{(D-1)}} + b_l^{kernel}\right) \tag{8b}$$

$$[R]_{U_D} = f\left(\sum_{z=1}^{C_0} Weight_l^{kernel,z} \star [R_E]_{F_{(D-1)}} + b_l^{kernel}\right) \tag{8c}$$

where $[R]_{U_D}$ denoting current R_U and $[R_E]_{F_D}$ representing current R_E, at D=1, 2, 3. Equation (8a) denoting a upsampling process of size $2 \times 2 \times 2$ on R_E. $[R_E]_{F(D-1)}$ representing the previous R_E. Equation (8b) representing a dense ASPP block of three parallel layers where layer l=1, 2, 3. Equation (8c) denoting a single convolution layer after the concatenation layer. Finally, the last R_U is reduced to match the brain MRIs' labels, followed by the softmax activation. The shape of the tensor after the softmax function is $3 \times 128 \times 128 \times 128$. After resampling, this shape is changed to the size of the original MRI for the submission purpose.

4 Experiments

We have used three benchmark datasets to validate our proposed architecture: BRATS 2018 and 2019 datasets [3–6, 35], and the six-month infant brain MRI iSeg-2019 dataset [44]. BRATS dataset is for brain tumor segmentation and the iSeg dataset for infant brain tissue segmentation. The difference for each subject among the BRATS and the iSeg datasets is the number of modalities; BRATS datasets have four different modalities, while the iSeg-2019 dataset has two different modalities. The description of BRATS datasets is discussed in the Section 4.1.1. The information of the iSeg-2019 dataset is given in the Section 4.2.1. The essential metrics used by the organizers in the MICCAI BRATS and iSeg competitions are explained in Section 4.4. Furthermore, we have worked on quantitative and qualitative analysis of BRATS and iSeg-2019 datasets.

4.1 Brain tumor segmentation challenge

4.1.1 Data description

BRATS 2018 and 2019 training sets contain 285 and 335 patients, respectively. According to the gliomas classification, both high-grade and low-grade patients are available in the BRATS training sets. We used 210 high-grade and 75 low-grade patients from the BRATS 2018 training dataset. In the meantime, 259 patients of high-grade and 76 patients of low-grade are selected from the BRATS 2019 training set. Each patient has four types of MRI: *native* (T1), *post-contrast T1-weighted* (T1ce), *T2-weighted* (T2) and *Fluid Attenuated Inversion Recovery* (FLAIR). The organizers performed different pre-processing steps on the entire data, such as skull-stripping, re-calculation to the equal 1mm^3 resolution, and all the scans of each case were co-registered to magnify the unhealthy tissues. The code of all pre-processing steps is now publicly available[1]. The manual segmentation of the entire training dataset was performed by the experts and provided by the organizers. $240 \times 240 \times 155$ is the dimension of each MRI modality. For each subject, the annotated labels has the values of 1 for the *necrosis and non-enhancing tumor* (NCR/NET), 2 for *peritumoral edema* (ED), 4 for *enhancing tumor* (ET), and 0 for background. The segmentation accuracy is measured by several metrics, where the predicted labels are evaluated by merging three regions, namely *whole tumor* (Whole Tumor or Whole: label 1, 2 and 4), *tumor core* (Tumor Core or Core: label 1 and 4), and *enhancing tumor* (Enhancing Tumor or Enhancing: label 1). We have evaluated our proposed model on validation datasets, 66 patients in the BRATS 2018, and 125 patients in BRATS 2019. Each patient in the validation datasets has no truth label.

4.2 Infant brain MRI segmentation challenge

4.2.1 Data description

In the MICCAI 2019 infant brain MRI competition, each team has access to three different datasets. 10 subjects are available in the iSeg-2019 training set. The validation dataset contains 13 subjects of one location, while 16 cases of three sites are available in the testing set. Each subject consists of two different MRI modalities: $T1$ and $T2$. The ground truth values are available with the training dataset. The dimension of each modality is $144 \times 192 \times 256$.

[1] https://cbica.github.io/CaPTk/preprocessing_brats.html

For each subject, the annotated labels has the values of 1 for the *cerebrospinal fluid* (CSF), 2 for *gray matter* (GM), 3 for *white matter* (WM), and 0 for background. At around 6 months of age, the intensity ranges of voxels in GM and WM in structural MRI images are largely overlapping (especially around the cortical regions), leading to the ambiguities creating the most challenge for tissue segmentation. The subjects of the validation and testing sets have no truth labels.

4.3 Implementation details

In this work, bias field correction and normalization steps are performed on each training dataset. During training, we performed the five-fold validation on each dataset. $4 \times 128 \times 128 \times 128$ is the input size for BRATS datasets while $2 \times 32 \times 32 \times 32$ is the size for the iSeg-2019 dataset. The batch size for the BRATS training datasets is 1, while the batch size is 4 for the iSeg-2019 dataset. We used Keras to build the proposed architecture. We have used Adam optimizer with 4×10^{-5} learning rate and a weight decay of 1×10^{-5}. We train our architecture for 60000 iterations with the BRATS training datasets. In contrast, 112800 iterations with the iSeg-2019 training dataset. Data augmentation have been undertaken on the fly for each patch, including flipping horizontally and rotating a random 90° to avoid the over-fitting problem during training. We used the Leaky ReLU non-linearity during training. We implemented instance normalization [20] because of a small batch size. The loss function is an important hyper-parameter during the training process. It helps balance the classes; in the BRATS and the iSeg training datasets, healthy tissues are bigger than unhealthy tissues. Different loss functions were previously proposed [14, 21, 24, 41]. We found that cross-entropy loss is not ideal with such kind of highly unbalanced datasets. Multi-label dice loss function has shown the remarkable results in highly imbalanced datasets [20, 36, 46]. We have used the loss function [36], while the number of samples per batch is one. (9) shows the mathematical representation of loss function.

$$Loss = -2 \sum_{d \in D} \frac{\sum_j pred_{j,d} truth_{j,d} + r}{\sum_j pred_{j,d} + \sum_j truth_{j,d} + r} \tag{9}$$

where $pred_{j,d}$ and $truth_{j,d}$ are the prediction obtained by softmax activation and ground truth at voxel j for class d, respectively. D is the total number of classes.

4.4 Evaluation metrics

For evaluating the BRATS and the iSeg datasets, we use the various metrics: the Dice Similarity Coefficient (DSC), the sensitivity, the specificity, the Hausdorff95 distance or modified Hausdorff distance (H95), and the average surface distance (ASD). Mathematically, each metric can be written as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{10}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$H95 = max\left\{ \max_{y \in G} d\left(y, S\right), \max_{y \in S} d\left(y, G\right) \right\} \tag{13}$$

$$ASD = \frac{1}{|G|} \sum_{x \in S} d\,(x, S) \tag{14}$$

where TP, FP, TN, and FN are the number of true positive, false positive, true negative, and false negatives voxels, respectively. For both of $H95$ and ASD, G and S are truth and segmented sets of voxels, respectively. For $H95$, $d(y, S)$ is point-to-set distance defined by: $d(y, S)$=$\min_{x \in s} \|y - x\|$, with $\|.\|$ denoting euclidean distance. We use the similar notation of $H95$ for ASD and $|.|$ denoting the cardinality of a set.

In the BRATS and the iSeg challenges, the ranking of the teams depends on the dice scores. We also report the best method based on the highest average dice scores in our work. DSC, sensitivity, and specificity evaluate the voxel-wise overlap between the truth and the segmented MRIs. H95 and ASD are the spatial distance-based metrics. The earlier is used for the BRATS datasets while the last one is used for the iSeg datasets. Furthermore, another name of metric H95 is modified hausdorff distance, commonly used in iSeg competitions.

5 Quantitative analysis

5.1 Brain tumor segmentation challenge

To check the capability of proposed architecture, we build four different architectures based on our proposed model. For the first architecture, Residual 3D U-Net, we replaced the residual-dilated blocks with residual blocks. Also, the dilated convolutional layers are replaced with non-dilated convolutions within each block. Moreover, the dense ASPP blocks are removed from the proposed architecture. For the second architecture, Residual-Dilated, we keep the residual-dilated blocks and remove the dense ASPP blocks from our proposed architecture. For the third architecture, Residual-Dense-Dilated, we implement the dense connections within each residual-dilated block and remove the proposed architecture's dense ASPP blocks. For the fourth architecture, Residual-Dilated-ASPP, we keep the residual-dilated blocks and remove the dense connections from the ASPP blocks. Table 1 depicts the details of the different models, including the proposed architecture. All the architectures are trained and validated with the BRATS 2018 datasets. To train each architecture, we used the BRATS datasets' training schemes, which were explained in the implementation details Section 4.3. The best fold of each architecture is used on the full training and the validation sets for the predicted MRIs. These predicted MRIs are then submitted to the organizers[2] for final scores. Each architecture's scores are based on several metrics: the Dice Similarity Coefficient, the sensitivity, the specificity, and the Hausdroff95 distances (H95). These metrics were explained in the evaluation metrics Section 4.4.

Table 2 depicts the results of all the models that include our proposed architecture. The number of parameters with the Residual-Dense-Dilated model is lowest compared to the algorithms: Residual 3D U-Net, and Residual-Dilated. We do not reach the Residual-Dense-Dilated model in terms of parameters with our proposed model and Residual-Dilated-ASPP architecture due to the presence of three parallel dilated convolutional layers in ASPP. Our proposed model's number of parameters is reduced compared to the Residual-Dilated-ASPP

[2] https://ipp.cbica.upenn.edu

Table 1 The architectural details

Architectures	Parameters	Residual	Dense	Rates at Encoder	Rates in ASPP	Channels
Residual 3D U-Net	4.00 M	At Encoder		1, 1		24
Residual-Dilated	4.00 M	At Encoder		1, 2		24
Residual-Dense-Dilated	2.90 M	At Encoder	Within Encoder	1, 2		24
Residual-Dilated-ASPP	5.40 M	At Encoder		1, 2	1, 2, 3, 5	24
Residual-Dilated Dense ASPP (Proposed)	4.53 M	At Encoder	Within ASPP	1, 2	1, 2, 3, 5	24

Each model is based on the combined approach of different methods

Table 2 The results of various architectures are based on different evaluation metrics

Architectures	Parameters	Metrics	Whole	Core	Enhancing
Residual 3D U-Net	4.00 M	DSC	92.293(90.664)	89.284(82.763)	76.141(74.899)
		Sensitivity	92.520(92.655)	90.464(85.474)	87.749(83.112)
		Specificity	99.514(99.456)	99.727(99.732)	99.797(99.748)
		H95(mm)	3.545(6.077)	3.937(10.105)	3.591(7.672)
Residual-Dilated	4.00 M	DSC	92.197(89.984)	89.160(83.063)	77.810(76.253)
		Sensitivity	91.976(89.921)	88.244(83.462)	83.790(80.968)
		Specificity	**99.615**(99.514)	99.769(99.804)	99.829(99.794)
		H95(mm)	3.619(5.748)	3.943(7.593)	3.500(3.710)
Residual-Dense-Dilated	2.90 M	DSC	92.874(90.638)	90.511(84.703)	78.279(76.663)
		Sensitivity	**93.429**(92.638)	**90.766**(85.587)	**91.976**(83.112)
		Specificity	99.486(99.429)	99.729(99.785)	99.814(99.750)
		H95(mm)	3.002(4.191)	3.406(7.202)	2.617(4.307)
Residual-Dilated-ASPP	5.40 M	DSC	92.999(89.852)	90.570(82.584)	79.014(77.621)
		Sensitivity	92.821(91.005)	89.489(83.976)	87.722(82.743)
		Specificity	99.459(99.462)	99.765(99.771)	99.830(99.780)
		H95(mm)	2.878(3.986)	**3.055**(7.704)	2.442(4.544)
(Proposed)	4.53 M	DSC	**93.294**(90.888)	**91.112**(84.463)	**80.904**(78.183)
		Sensitivity	93.377(92.660)	89.982(85.474)	87.857(83.608)
		Specificity	99.536(99.429)	**99.805**(99.785)	**99.851**(99.750)
		H95(mm)	**2.710**(5.748)	3.457(7.202)	**2.053**(4.307)

The mean scores of each metric of the training dataset are shown without brackets, while brackets have the validation scores. The best scores of the training and validation datasets are respectively highlights in bold and blue

architecture by introducing the dense connections with a growth rate of 12 in ASPP. For the Residual-Dilated model, the dice scores of the three types of cancer tumors increase (on validation dataset) compared to the Residual 3D U-Net model. The increased dice scores reaffirm the potential of the dilation blocks with residual connections. For Residual-Dilated-ASPP and Residual-Dense-Dilated models, the whole tumor's scores for the sensitivity and the H95 distances increase, and the specificity decreases; thereby, the occurrence of the false positives increases.

The Residual-Dilated model compared to the models: Residual-Dense-Dilated and Residual-Dilated-ASPP, the sensitivity scores and the H95 distances of the whole tumor decrease, and the specificity increases; thereby, the occurrence of false negatives increases. Our proposed model can balance the events of false positives and false negatives based on the combined approach of the residual network, dilation, and dense ASPP. For the Residual-Dilated, Residual-Dense-Dilated, and Residual-Dilated-ASPP models, the tumor core's dice scores and enhancing tumor increases. Also, the sensitivity and the H95 distances decrease for the tumor core and enhance tumor but increase the specificity. Thus, the aggregation process improves the dice scores of the tumor core and the enhancing tumor. In comparing our proposed architecture, the Residual $3D$ U-Net model obtained the Dice similarity coefficient's low scores, a deciding metric for the best methods in BRATS competitions. The combined approach of the residual connections, dilation, and ASPP gives excellent results, as shown in Table 2 in terms of parameters.

5.1.1 Comparison with the best methods

Table 3 shows the comparison of our proposed work with the state-of-art methods of the BRATS 2018 validation dataset [2, 7, 8, 10, 14, 17, 18, 25, 27, 32, 33, 37, 38, 43, 45, 47, 49]. Here, we compared the results based on the dice scores. The proposed architecture secures best average scores against all the other algorithms, even obtained the higher scores than the ensembling of several architectures [2, 25, 27]. Furthermore, our work can save the time which is spent in performing the complex post-processing *Conditional Random Field* (CRF) Chandra [ResNet + CRF, V-Net + CRF] et al. [8] and test-time augmentations (TTA) Wang [3D UNer + TTA, Multi-class WNet + TTA] et al. [47], common strategies for removing the false-positive voxels. Based on the higher dice scores, our proposed model is more generalized on the unseen validation dataset.

We choose three different algorithms from Table 3 to justify our proposed approach of the residual connections, dilation, and dense ASPP. These three algorithms are: Chen et al. [10], Sun [DFKZ Net] et al. [43], and Chandra et al. [8]. Chen et al. [10] performed the division operation on a $3D$ weighted layer within each block of a residual network; the resulting layers were increased the number of parameters. Our proposed architecture scores exhibit the necessity of a $3D$ Convolutional layer to process the $3D$ brain MRIs. Sun [DFKZ Net] et al. [43] used a residual-based 3D U-Net model of Isensee et al. [20] with BRATS 2018 datasets; non-dilated convolution layers were used in the first part of the architecture. To reaffirm the potential of enhanced sizes with the residual part, we removed the dense ASPP blocks from our proposed architecture; the resulting architecture became a Residual-Dilated model (see row number 2 of Table 2). Its scores reaffirm our contribution of implementing the different dilation rates to preserve more information about the tumor's small sizes. Chandra et al. [8] enhances the sizes of the receptive field to extract the complete information of an image [31] just before the softmax activation. For this, an atrous spatial pyramid pooling (ASPP) was used. The number of parameters was increased as three residual U-Nets

Table 3 Performance evaluation of different architectures on the validation dataset (BRATS 2018)

Architectures	Whole tumor	Tumor core	Enhancing tumor
Albiol [Ensemble] et al. [2]	88.10	77.70	77.30
Carver et al. [7]	88.00	77.00	71.00
Chen et al. [10]	89.35	83.09	74.93
Chandra [ResNet] et al. [8]	86.80	80.10	74.00
Chandra [ResNet + CRF] et al. [8]	87.20	79.90	74.10
Chandra [V-Net] et al. [8]	89.90	81.00	76.60
Chandra [V-Net + CRF] et al. [8]	90.10	81.30	74.93
Feng [Model 1] et al. [14]	90.15	82.37	76.88
Feng [Model 2] et al. [14]	90.66	82.48	76.77
Feng [Model 3] et al. [14]	90.40	83.06	76.95
Feng [Model 4] et al. [14]	89.90	81.04	77.07
Feng [Model 6] et al. [14]	89.17	81.49	76.16
Hu et al. [17]	88.00	74.00	69.00
Hua et al. [18]	90.48	83.64	77.68
Kao [Average scores of three 3D U-Net models] et al. [25]	89.40	77.50	76.40
Kori [3D Dense U-Net (Model 1)] et al. [27]	85.00	74.00	71.00
Kori [2D Dense U-Net (Model 2] et al. [27]	87.00	73.00	71.00
Kori [3D Hierarchical Architecture (Model 3] et al. [27]	85.00	73.00	71.00
Kori [Ensembling (Model 1 + Model 2 + Model 3)] et al. [27]	89.00	76.00	76.00
Ma [Complementary fusion] et al. [32]	87.20	77.30	74.30
Ma [Ordinary fusion] et al. [32]	85.10	75.10	70.90
Marcinkiewicz et al. [33]	89.80	81.18	75.19
Nuechterlein et al. [37]	85.50	78.20	66.50
Rezaei et al. [38]	84.00	79.00	63.00
Sun [DFKZ Net] et al. [43]	89.31	82.46	76.77
Tuan et al. [45]	81.87	69.98	68.25
Wang [3D UNet] et al. [47]	86.38	76.58	73.44
Wang [3D UNet + TTA] et al. [47]	87.31	78.32	75.43
Wang [Multi-class WNet] et al. [47]	89.98	72.53	75.70
Wang [Multi-class WNet + TTA] et al. [47]	89.56	73.04	77.07
Weninger et al. [49]	88.90	75.80	71.20
RD^2A (Proposed)	**90.88**	**84.46**	**78.18**

For comparison, only DSC scores are shown. The best scores are highlights in bold

Table 4 The results of the proposed architecture are based on different evaluation metrics

Dataset	Metrics	Whole	Core	Enhancing
BRATS 2019 Training	DSC	92.533	90.295	78.108
	Sensitivity	93.576	90.855	87.207
	Specificity	99.466	99.729	99.830
	H95(mm)	3.480	3.522	3.026
BRATS 2019 Validation	DSC	90.357	82.348	71.934
	Sensitivity	91.578	80.807	80.098
	Specificity	99.446	99.781	99.819
	H95(mm)	4.690	7.088	3.316

The mean score of each metric is obtained after evaluating the predicted MRIs (of the BRATS 2019 training and validation datasets) via the online web portal https://ipp.cbica.upenn.edu. Our team name is Tyagi for the BRATS 2019 datasets in the competition

were used. Moreover, Chandra et al. [8] used the big rates of dilation, thereby introducing the gridding problem; the vital local information was lost. Our proposed architecture scores with only 4.53 M of parameters exhibit the combined residual-dense connections' necessity.

In summary, our proposed architecture achieved excellent results in all types of tumors. Furthermore, all the predicted MRIs of the BRATS 2019 datasets were submitted to the organizer's webpage[3] for the online evaluation. The evaluation scores are shown in Table 4.

5.2 Infant brain MRI segmentation challenge

To ensure the proposed architecture's capability, we also validate our architecture on the iSeg-2019 datasets. The training dataset (10 subjects) is divided into five-folds. In each fold, 8 subjects are selected for the training and remaining for the validation. We have chosen the best fold to evaluate the iSeg-2019 validation (13 subjects) and the testing datasets (16 subjects). Table 5 shows the results of all methods for iSeg-2019 validation and testing datasets. The scores without brackets are related to the validation dataset while the remaining to the iSeg-2019 testing dataset. The average score of each metric, especially DSC, is best with our proposed model than the MASI (baseline), long, and UBC001 methods. In the meantime, the CSF average scores for the metrics DSC and the ASD are lower with our presented work compared to the lyh and the tiantian methods. Furthermore, the scores of the Brain_Tech method are higher than our proposed work. In short, the Brain_Tech method is best with the iSeg-2019 validation dataset. At the same time, we secure the best scores on the testing dataset than the validation dataset's top method. The validation dataset subjects belong to only one site, while the 16 subjects of the testing dataset are collected from three different locations. Our proposed model can perform better generalization on the unseen dataset of two or more sites based on the higher testing scores. Furthermore, the best testing dice scores are successfully distinguishing the contrast between the *gray matter* (GM) and *white matter* (WM) tissues.

[3] https://ipp.cbica.upenn.edu/

Table 5 The results of our proposed architecture, along with all methods

Methods	Metrics	CSF	GM	WM
MASI(baseline)	DSC	67.100(67.000)	53.600(60.700)	64.500(68.100)
	Hausdorff95(mm)	14.405(18.013)	30.831(20.375)	14.695(10.153)
	ASD	0.912(1.332)	1.310(1.033)	1.561(1.308)
long	DSC	89.800(78.80)	84.900(73.500)	81.700(76.200)
	H95(mm)	10.369(9.503)	7.595(9.140)	8.335(9.853)
	ASD	0.290(**0.565**)	0.548(0.663)	0.636(0.828)
lyh	DSC	92.200(79.200)	86.700(70.500)	84.130(69.900)
	H95(mm)	12.477(62.194)	22.807(18.795)	11.666(13.660)
	ASD	0.211(0.632)	0.501(0.775)	0.583(1.110)
Brain_Tech	DSC	96.100(79.500)	92.800(69.400)	91.100(78.000)
	H95(mm)	8.873(11.421)	5.724(9.516)	7.114(9.237)
	ASD	0.108(0.626)	0.300(0.735)	0.347(0.886)
UBC001	DSC	87.700(69.900)	86.800(66.000)	81.900(63.500)
	H95(mm)	9.370(9.806)	9.736(9.886)	8.964(13.865)
	ASD	0.333(0.838)	0.547(0.961)	0.689(1.693)
tiantian	DSC	91.200(79.600)	85.400(71.300)	82.500(64.200)
	H95(mm)	10.760(101.137)	10.698(20.358)	9.414(16.117)
	ASD	0.237(0.862)	0.530(0.802)	0.628(1.288)
(Proposed)	DSC	90.020(**79.804**)	86.900(**77.925**)	84.800(**80.569**)
	H95(mm)	9.463(11.626)	6.339(**8.131**)	7.111(**8.752**)
	ASD	0.263(0.611)	0.509(**0.655**)	0.576(**0.735**)

These models are evaluated on 13 and 16 subjects of the iSeg-2019 validation and testing datasets. The mean scores of each metric of the validation dataset are shown without brackets, while brackets have the testing scores. Our team name for the MICCAI iSeg-2019 competition is Legand. The best scores of the validation and testing datasets are respectively highlights in blue and bold. The scores can be accessed via http://iseg2019.web.unc.edu/evaluation-results/

6 Qualitative analysis

6.1 Brain tumor segmentation

The segmentation results of our proposed architecture are shown in Fig. 8. We choose two different patients from the BRATS 2018 training dataset. For these two patients, we only visualize the T1ce modality with ground-truth and prediction. Figure 8a and b represent the ground-truth and prediction with the T1ce modality of one patient, respectively. Figure 8c and d of another patient represents the T1ce modality with the ground-truth and prediction, respectively. Moreover, Fig. 8a and c exhibit the variations among the shape, size, and location of the tumors in different patients. As depicted by the predicted T1ce modality in Fig. 8d, our proposed algorithm has the potential to segment the big size of the whole tumor and the small size of the tumor core.

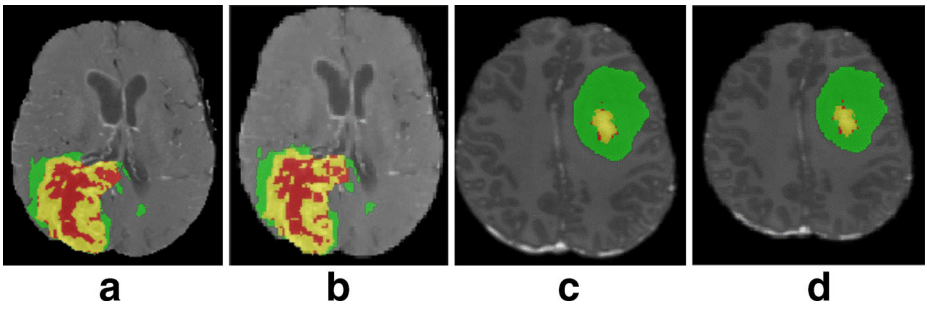

Fig. 8 Segmentation results using our proposed architecture. **a** and **b** represent the ground-truth and prediction on T1ce modality, respectively. **c** and **d** denote the ground-truth and prediction on the T1ce modality of another patient, respectively. Each color represents a different tumor: red for Tumor core, green for the Whole tumor, and yellow for Enhancing tumor

Our proposed algorithm failed for some patients of the BRATS 2018 training dataset. A case of wrong segmentation is depicted in Fig. 9. We only visualize the $T1ce$ modality with ground-truth and the prediction. A long orange arrow in Fig. 9b exhibits the instance of the wrong segmentation, in which our proposed algorithm wrongly predicted the background label as a tumor core. We will investigate to solve the wrong prediction in the future work through the combined loss functions.

6.2 Infant brain MRI segmentation

Figure 10 shows the segmentation results of our proposed architecture. We choose a subject from the iSeg-2019 training dataset. We demonstrated ground-truth and prediction of the selected subject on a $T1$ modality. Figure 10a and b represent the ground-truth and prediction, respectively. Predicted visualization exhibits the potential of the proposed algorithm for infant brain MRI segmentation.

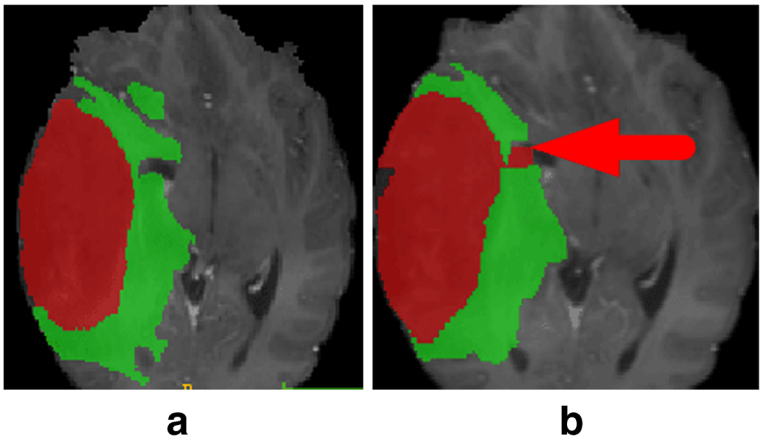

Fig. 9 An instance of wrong segmentation results. **a** and **b** represent the ground-truth and prediction on $T1ce$ modality of a patient, respectively. **b** represents an instance of the wrong segmentation; our proposed algorithm is wrongly predicted the background label as a tumor core that is shown by a long orange arrow

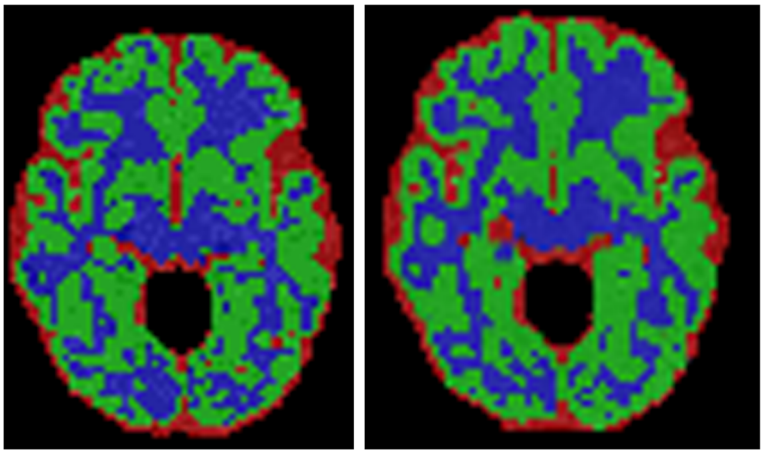

Fig. 10 Training segmentation results using our proposed architecture. **a** and **b** represent the ground-truth and prediction on T1 modality, respectively. Different colors represent brain tissues: red for CSF, green for GM, and blue for WM

7 Discussion and conclusion

We have proposed a model with the combination of the residual connections, dilation, and dense ASPP. Different atrous rates are chosen in the residual-dilation blocks to avoid the gridding problem. In the meantime, dense connections are employed to reduce the parameters. Dense ASPP blocks exploit the multi-scale information to avoid the ambiguities among brain MRIs' labels and tissues. The multi-label cost function is used to prevent the imbalanced data problem. Augmentation techniques such as flipping and rotations are used to avoid the over-fitting problem during training. Finally, the combined approach achieved outstanding results with different brain MRI datasets.

We cannot train our network on big patch sizes due to memory limitations, especially with the BRATS datasets. Chen et al. [9] implemented ASPP on a big patch size with improved results. In the future, we will try our proposed approach on multiple medical imaging problems, especially for the kidney tumor segmentation using big patch sizes. Kidney tumor segmentation is a very challenging problem due to lack of information from only one modality in the MICCAI KiTS 2019 dataset [16]. In our work, we experimentally proved that the parameters could be efficiently reduced with improved results. Our proposed architecture has the potential to solve the problem of other medical imaging tasks. Furthermore, we will propose an architecture with weighted majority schemes and the study on the different normalization layers with varying batch sizes.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant No.61672250 and the Hubei Provincial Development and Reform Commission Project in China.

Availability of data and material

BRATS 2018 datasets The description of the datasets and procedures to download them can be accessed[4].

BRATS 2019 datasets The description of the datasets and procedures to download them can be accessed[5].

iSeg-2019 datasets The description of the datasets and procedures to download them can be accessed[6].

Declarations

Conflicts of interest/Competing interests The authors declare that there is no conflict of interest/competing interests.

References

1. Ahmad P, Qamar S, Hashemi SR, Shen L (2020) Hybrid labels for brain tumor segmentation. In: Crimi A, Bakas S (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 158–166
2. Albiol A, Albiol A, Albiol F (2019) Extending 2D deep learning architectures to 3D image segmentation problems. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 73–82
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C (2017) Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C (2017) Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive 286
5. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann JB, Farahani K, Davatzikos C (2017) Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data 4:170117. https://doi.org/10.1038/sdata.2017.117
6. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara RT, Berger C, Ha SM, Rozycki M, Prastawa M, Alberts E, Lipková J, Freymann JB, Kirby J, Bilello M, Fathallah-Shaykh HM, Wiest R, Kirschke J, Wiestler B, Colen RR, Kotrotsou A, LaMontagne P, Marcus DS, Milchenko M, Nazeri A, Weber M, Mahajan A, Baid U, Kwon D, Agarwal M, Alam M, Albiol A, Albiol A, Varghese A, Tuan TA, Arbel T, Avery ABP, Banerjee S, Batchelder T, Batmanghelich K, Battistella E, Bendszus M, Benson E, Bernal J, Biros G, Cabezas M, Chandra S, Chang YJ et al (2018) Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the {BRATS} Challenge. CoRR arXiv:1811.02629
7. Carver E, Liu C, Zong W, Dai Z, Snyder JM, Lee J, Wen N (2019) Automatic brain tumor segmentation and overall survival prediction using machine learning algorithms. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 406–418
8. Chandra S, Vakalopoulou M, Fidon L, Battistella E, Estienne T, Sun R, Robert C, Deutsch E, Paragios N (2019) Context Aware 3D CNNs for Brain Tumor Segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 299–310
9. Chen LC, Papandreou G, Schroff F, Adam H (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. CoRR arXiv:1706.05587

[4] https://www.med.upenn.edu/sbia/BraTS2018/data.html

[5] https://www.med.upenn.edu/cbica/BraTS2019/data.html

[6] https://iseg2019.web.unc.edu/data/

10. Chen W, Liu B, Peng S, Sun J, Qiao X (2019) S3d-UNet: Separable 3D U-Net for Brain Tumor Segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 358–368
11. Devalla SK, Renukanand PK, Sreedhar BK, Perera SA, Mari JM, Chin KS, Tun TA, Strouthidis NG, Aung T, Thiery AH, Girard MJA (2018) {DRUNET:} {A} Dilated-Residual U-Net Deep Learning Network to Digitally Stain Optic Nerve Head Tissues in Optical Coherence Tomography Images. CoRR arXiv:1803.00232
12. Dolz J, Ayed IB, Desrosiers C (2018) Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities. CoRR arXiv:1810.07003
13. Dolz J, Xu X, Rony J, Yuan J, Liu Y, Granger E, Desrosiers C, Zhang X, Ayed IB, Lu H (2018) Multi-region segmentation of bladder cancer structures in {MRI} with progressive dilated convolutional networks. CoRR arXiv:1805.10720
14. Feng X, Tustison N, Meyer C (2019) Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 279–288
15. He K, Zhang X, Ren S, Sun J (2015) Deep Residual Learning for Image Recognition. CoRR arXiv:1512.03385
16. Heller N, Sathianathen N, Kalapara A, Walczak E, Moore K, Kaluzniak H, Rosenberg J, Blake P, Rengel Z, Oestreich M et al (2019) The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv:1904.00445
17. Hu Y, Liu X, Wen X, Niu C, Xia Y (2019) Brain Tumor Segmentation on Multimodal MR Imaging Using Multi-level Upsampling in Decoder. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 168–177
18. Hua R, Huo Q, Gao Y, Sun Y, Shi F (2019) Multimodal brain tumor segmentation using cascaded V-Nets. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 49–60
19. Huang G, Liu Z, Weinberger KQ (2016) Densely Connected Convolutional Networks. CoRR arXiv:1608.06993
20. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2018) Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the {BRATS} 2017 Challenge. CoRR arXiv:1802.10508
21. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2019) No New-Net. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, pp 234–244
22. Jégou S, Drozdzal M, Vázquez D, Romero A, Bengio Y (2016) The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. CoRR arXiv:1611.09326
23. Kamnitsas K, Bai W, Ferrante E, McDonagh SG, Sinclair M, Pawlowski N, Rajchl M, Lee MCH, Kainz B, Rueckert D, Glocker B (2017) Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. CoRR arXiv:1711.01468
24. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 36:61–78
25. Kao PY, Ngo T, Zhang A, Chen JW, Manjunath BS (2019) Brain tumor segmentation and tractographic feature extraction from structural MR images for overall survival prediction. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 128–141
26. Kermi A, Mahmoudi I, Khadir MT (2019) Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 37–48
27. Kori A, Soni M, Pranjal B, Khened M, Alex V, Krishnamurthi G (2019) Ensemble of fully convolutional neural network for brain tumor segmentation from magnetic resonance images. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 485–496
28. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

29. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J (2020) A survey on U-shaped networks in medical image segmentations. Neurocomputing 409:244–258. https://doi.org/10.1016/j.neucom.2020.05.070. http://www.sciencedirect.com/science/article/pii/S0925231220309218
30. Long J, Shelhamer E, Darrell T (2014) Fully Convolutional Networks for Semantic Segmentation. CoRR arXiv:1411.4038
31. Luo W, Li Y, Urtasun R, Zemel RS (2017) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. CoRR arXiv:1701.04128
32. Ma J, Yang X (2019) Automatic Brain Tumor Segmentation by Exploring the Multi-modality Complementary Information and Cascaded 3D Lightweight CNNs. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 25–36
33. Marcinkiewicz M, Nalepa J, Lorenzo PR, Dudzik W, Mrukwa G (2019) Segmenting Brain Tumors from MRI Using Cascaded Multi-modal U-Nets. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 13–24
34. Mehta R, Arbel T (2019) 3D U-Net for brain tumour segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 254–266
35. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber M, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Criminisi A (2015) The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 34(10):1993–2024. https://doi.org/10.1109/TMI.2014.2377694
36. Milletari F, Navab N, Ahmadi SA (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. CoRR arXiv:1606.04797
37. Nuechterlein N, Mehta S (2019) 3d-ESPNet with Pyramidal Refinement for Volumetric Brain Tumor Image Segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 245–253
38. Rezaei M, Yang H, Meinel C (2019) voxel-GAN: Adversarial Framework for Learning Imbalanced Brain Tumor Segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 321–333
39. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR arXiv:1505.04597
40. Roy Choudhury A, Vanguri R, Jambawalikar SR, Kumar P (2019) Segmentation of Brain Tumors Using DeepLabv3+. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 154–167
41. Samanta A, Saha A, Satapathy SC, Fernandes SL, Zhang YD (2020) Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recogn Lett 135:293–298. https://doi.org/10.1016/j.patrec.2020.04.026. http://www.sciencedirect.com/science/article/pii/S0167865520301483
42. Sarker MMK, Rashwan HA, Akram F, Banu SF, Saleh A, Singh VK, Chowdhury FUH, Abdulwahab S, Romani S, Radeva P, Puig D (2018) SLSDeep: Skin Lesion Segmentation Based on Dilated Residual and Pyramid Pooling Networks. CoRR arXiv:1805.10241
43. Sun L, Zhang S, Luo L (2019) Tumor segmentation and survival prediction in glioma with deep learning. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 83–93
44. Sun Y, Gao K, Wu Z, Lei Z, Wei Y, Ma J, Yang X, Feng X, Zhao L, Phan TL, Shin J, Zhong T, Zhang Y, Yu L, Li C, Basnet R, Ahmad MO, Swamy MNS, Ma W, Dou Q, Bui TD, Noguera CB, Landman B, Gotlib IH, Humphreys KL, Shultz S, Li L, Niu S, Lin W, Jewells V, Li G, Shen D, Wang L (2020) Multi-Site Infant Brain Segmentation Algorithms: The iSeg-2019 Challenge
45. Tuan TA, Tuan TA, Bao PT (2019) Brain Tumor Segmentation Using Bit-plane and UNET. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 466–475
46. Wang G, Li W, Ourselin S, Vercauteren T (2017) Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks. CoRR arXiv:1709.00382
47. Wang G, Li W, Ourselin S, Vercauteren T (2019) Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In: Lecture Notes in Computer Science (including

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 11384 LNCS. Springer, pp 61–72. https://doi.org/10.1007/978-3-030-11726-9_6
48. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell GW (2017) Understanding Convolution for Semantic Segmentation. CoRR arXiv:1702.08502
49. Weninger L, Rippel O, Koppers S, Merhof D (2019) Segmentation of Brain Tumors and Patient Survival Prediction: Methods for the braTS 2018 Challenge. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 3–12
50. Xie S, Sun C, Huang J, Tu Z, Murphy K (2017) Rethinking Spatiotemporal Feature Learning For Video Understanding. CoRR arXiv:1712.04851
51. Xu Y, Gong M, Fu H, Tao D, Zhang K, Batmanghelich K (2019) Multi-scale Masked 3-D U-Net for Brain Tumor Segmentation. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T (eds) Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp 222–233
52. Yang M, Yu K, Zhang C, Li Z, Yang K (2018) DenseASPP for Semantic Segmentation in Street Scenes. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 3684–3692
53. Yu F, Koltun V (2016) Multi-scale Context Aggregation by Dilated Convolutions. coRR arXiv:1511.0
54. Yu F, Koltun V, Funkhouser TA (2017) Dilated Residual Networks. CoRR arXiv:1705.09914
55. Zhang J, Jin Y, Xu J, Xu X, Zhang Y (2018) MDU-Net: Multi-scale Densely Connected U-Net for biomedical image segmentation. CoRR arXiv:1812.00352
56. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: {A} Nested U-Net Architecture for Medical Image Segmentation. CoRR arXiv:1807.10165