Noname manuscript No.
(will be inserted by the editor)

Dilated Convolutions based 3D U-Net for Multi-Modal Brain Image Segmentation

Ouissam Kemassi [1] · Oussama Maamri[1] · Khadra Bouanane [1] · Kriker Ouissal [1]

[1]Université Kasdi Merbah Ouargla (FNTIC)

Received: 2021

Abstract Several deep learning based medical image segmentation methods use U-Net architecture and its variants as a baseline model. This is because U-Net has been successfully applied to many other tasks. it was noticed that the U-Net-based models are unable to extract features for segmenting small masks or fine edges.
To overcome this issue, we propose a new 3D U-Net-based model, baptized Y- Net. In this model, we make use of dilated convolution which has shown its effectiveness in grasping different features at different scales. This allows us to capture more information from small anatomical parts.
Our model is assessed on MRbrains13 dataset for brain tissue segmentation task. Compared to the traditional UNet 3D, the obtained results show that the proposed model performs well, especially in segmenting white matter and grey matter tissues.

Keywords Brain Image Segmentation · UNet 3D · MRI Modalities · Dilated Convolutions.

1 Introduction

In the field of medical image analysis, image segmentation is considered as the most challenging task that aims to identify the pixels of organs or lesions from background medical images produced from the process of medical imaging.
Image segmentation is frequently utilized in brain MRI analysis for measuring and

O.Kemassi
E-mail: kmasyw@gmail.com

O.Maamri
E-mail: maamri95@gmail.com

Kh.Bouanane
E-mail: khadra.bouanane@gmail.com

O.Kriker
E-mail: Krikerouissal@gmail.com

visualizing brain structures, assessing brain changes, identifying diseased regions, surgical planning, and image-guided therapies.
Due to its importance, the automation of this task has been extensively studied over decades, where researchers aim to improve the segmentation of medical images using new methods.
One of the most successful approaches that have emerged this last decade is deep learning-based models that have shown a significant improvement in segmentation accuracy.
One of the most popular deep learning models for medical image segmentation is U-Net [1,2] which has shown to be a very effective model. In particular, it was recorded in many works that U-Net based models achieve high performance in segmenting medical images [3–5].
Its success is attributed to skip connections between the encoder and decoder parts of the architecture.
However, U-Net based models lack focus in extracting features for segmentation of small structures. It was observed that this class of networks perform poorly in detecting smaller structures and are thus unable to segment boundary regions properly [6].
To overcome this issue, we propose a new 3D UNet based model, baptized YNet. In this model, we make use of dilated convolution which has shown its effectiveness in grasping features at different scales [7]. This allows to capture more information from small anatomical parts and thus, enhance the performance of the model.
In the remainder of this paper we first present in Section 2, some recent U-Net based models that were proposed to tackle brain image segmentation. We give a detailed description of the proposed model in Section nd in Section 4, experimental results are provided.

2 Related Work

Based on U-Net architecture, several previous works have proposed efficient models for brain image segmentation task.
Rehman et al. [8] proposed a 2D image segmentation method, called BU-Net, to contribute in brain tumor segmentation research, where residual extended skip (RES) and wide context (WC) are used along with the customized loss function in the baseline U-Net architecture. These modifications allow finding more differing features to get better segmentation performance.
Ozgun et al. [2], proposed a model for volumetric segmentation that learns from sparsely annotated volumetric images. This network extends the previous U-Net architecture from Ronneberger et al.(reference) by replacing all 2D operations with their 3D counterparts.
To tackle the poor performance of the U-net architecture when segmenting small structures, Valanarasu et al. [6] proposed Kiunet, an overcompleted convolutional network, which achieves an improved performance comparing to all the recent methods with an additional benefit of fast convergence.
KiU-Net and KiU-Net3D were applied on five different datasets covering various image modalities. Brain MRI images from BraTS2020 Challenge were used to assess KiU-Net3D.
Chen et al. [9] proposed a novel separable 3D U-Net architecture using separable

3D convolutions. This method achieved a mean enhancing tumor, whole tumor, and tumor core Dice scores in Preliminary results on BraTS 2018 validation set. This novel separable 3D U-Net architecture succeeded to overcome the issue of 2D convolutions, which cannot make a full use of the spatial information of volumetric medical images, by using 3D convolutions in such a way to prevent the high expensive computational cost and memory demand.

Peng et al. [10] proposed a Multi-Scale 3D U-Net architecture, which uses several U-Net blocks to capture long distance spatial information at different resolutions in order to extract and utilize sufficient features.

Feng et al. [11] proposed to use a 3D U-Net structure on extracted 2D/3D patches to predict the class labels for all pixels in the patch, in order to improve the performance of U-Net for brain tumor images.

On the other hand, we can notice that many additional operations have been applied in order to enhance the performance of the proposed models. Among these operations, dilated convolution has shown its ability and potential in improving segmentation results. The advantage in using dilated convolution in image segmentation has been highlighted in [7], where it was shown that dilated convolution permits the increase of the receptive fields without losing resolution or coverage. This makes the model more efficient in capturing different features at different scales. Therefore, it can be very beneficial for segmenting small anatomical parts in biomedical domain [12,13].

In literature, dilated convolution has been used to improve the segmentation accuracy. For instance, Zhang et al. [12] proposed a new end-to-end network based on ResNet and U-Net, in which the usual pooling layer is replaced with convolutional layer which permits to reduce the information loss extent.

Lei et al. [13] proposed a dual aggregation network to adaptively aggregate different information from infant's brain MRI images. Based on 3D U-Net, authors used the dilated convolution pyramid down-sampling module to solve the problem of loss of spatial information on the down-sampling process. This model achieved the first place in Iseg-2019 challenge.

Li et al. [14] proposed a novel model called Multi modal aggregation Network (MMAN). This model is an innovative deep network that was designed to extract and aggregate multiscale features of brain tissues from multi-modality MR images in order to get a precise and perfect segmentation. This network is based on an inception model with a variable dilation.

In this paper, we propose a novel model, baptized Y-Net. The main purpose is to make use of dilated convolution benefits to enhance a U-Net architecture. the model is detailed in the following section.

3 Description of our model

3.1 Network Architecture

Our proposed model contains three components: a regular encoder, a dilated encoder and a decoder.

– The regular encoder comprises four downsampling layers. Each layer uses two convolutions "conv3D " with a kernel size of $3 \times 3 \times 3$ voxels per block. The rectified linear units (ReLUs) is used as an activation function and a subsequent

3D-max pooling operation "MaxPool3D" is then performed.

- The dilated encoder includes four downsampling layers, where the usual convolution operation is replaced by dilated convolution with a predefined rate. This allows our model to enlarge the receptive field and thus capture fine details and accurate edges in the image.
  In this encoder, each layer uses two convolutions "conv3D "with a kernel size of $3 \times 3 \times 3$ voxels per block and a dilation-rate size of $2 \times 2 \times 2$. A rectified linear units (ReLUs) is used as an activation function.
- The output features of the regular and the dilated encoders are aggregated via an additional block, then the result of this operation is inputted to the decoder component.
- The decoder comprises four upsampling convolution layers where each layer uses two convolutions with a kernel size of $3{\times}3{\times}3$ voxels per block, the rectified linear units (ReLUs) as activation functions.

The model architecture is illustrated in Figure 1.

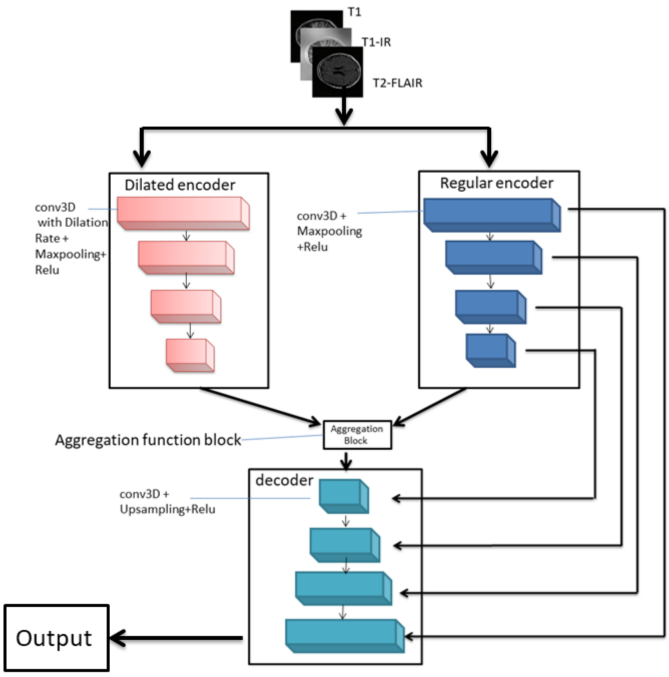

Fig. 1 Architecture of Y-Net

3.2 Training

The enhancement of training phase of our model is due to the aggregation of the different learned features that are provided by regular and dilated encoders.

This improvement in efficiency is due to dilated convolution that allows us to deal with object dependencies on different scales without reducing the image resolution. The model is trained by minimizing the categorical cross entropy between prediction and label.

$$L = -\sum_{i \in \Omega} \sum_{c} y_{ic} \log p_{ic} + (1 - y_{ic}) \log(1 - p_{ic}) \tag{1}$$

where

- p_{ic} is the probability that the $i$the voxel belongs to the c-th class.
- y_{ic} is the corresponding ground truth.
- Ω denotes all pixels in predicted segmentation result p.

4 Experimental Results

In this experiment, we assess the performance of Y-Net compared to the U-Net 3D [2] for the task of brain tissue segmentation from MRbrainS13 dataset.

4.1 Data

The MRBrainS13 dataset contains 20 subjects where 15 are used for testing and 5 for training. Each subject possesses three modalities: T1-weighted scan, T1-weighted IR scan and T2-FLAIR scan. The size of those modalities is $240 \times 240 \times 48$. The segmented tissues are labeled as follows:

0: Background (everything outside the brain)
1: Cerebrospinal fluid (including ventricles)
2: Gray matter (cortical gray matter and basal ganglia)
3: White matter (including white matter lesions)

We train the models by feeding them with extracted sub-volumes of size 64×64×16. By using the sub-volume method, we increase the number of training examples and reduce the memory requirement.

4.2 Evaluation Metrics

To measure the performance of the two models, we use the Dice Coefficient(DC), which is one of the most popular evaluation segmentation performance metrics. Dice Coefficient is a similarity measurement technique based on a statistical approach. It is defined by the overlap region between the predicted segmentation results and ground truth, its result is expressed as a percentage that varies from 0 % (total mismatch) to 100 % (perfect match) [15]. Formally, the DC formula is expressed as follows:

$$DC(G,S) = 2\frac{G \cap S}{G + S}.100\%$$

.

4.3 Results

Results of the training of Y-Net and U-Net 3D are summarzied in Table 1, As we can notice, U-Net 3D outperforms Y-Net in the task of segmentation of cerebrospinal fluid tissue while Y-Net achieves better results on on the Gray Matter(GM) and White Matter(WM) tissues.

Table 1 Comparison of dice score for brain tissue segmentation (MRbrainS13 dataset).

	Cerebrospinal fluid (CSF)	Gray matter (GM)	White matter (WM)
Y-Net	63.39	72.35	80.0
U-Net 3D	65.05	69.36	78.66

5 Conclusion

In this paper, we presented a new U-Net based model for brain tissue segmentation. Y-NET aims to overcome the inability of U-Net models in extracting features for segmentation of small structures. To do so, we proposed to exploit the benefits of dilated convolution by integrating a dilated encoder to the usual architecture. Information from different convolution blocks is then aggregated and used to extract meaningful features at different scales.
Experimental results show the effectiveness of the proposed model for the segmentation of white matter and gray matter tissues, compared to U-NET 3D, on the MRbrains13 dataset.
However, In order to corroborate the model, further experiments must be conducted for brain tissue segmentation task and other tasks as well. Especially that the model shows less performance compared to U-Net3D, in segmenting the cerebrospinal fluid tissue.

References

1. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
2. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International conference on medical image computing and computer-assisted intervention, Springer, 2016, pp. 424–432.
3. G. Zeng, G. Zheng, Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 136–140.
4. B. Yang, W. Zhang, Fd-fcn: 3d fully dense and fully convolutional network for semantic segmentation of brain anatomy, arXiv preprint arXiv:1907.09194 (2019).
5. J. Chang, X. Zhang, M. Ye, D. Huang, P. Wang, C. Yao, Brain tumor segmentation based on 3d unet with multi-class focal loss, in: 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2018, pp. 1–5.

6. J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, V. M. Patel, Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation, arXiv preprint arXiv:2010.01663 (2020).
7. R. Hamaguchi, A. Fujita, K. Nemoto, T. Imaizumi, S. Hikosaka, Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery, in: 2018 IEEE winter conference on applications of computer vision (WACV), IEEE, 2018, pp. 1442–1450.
8. M. U. Rehman, S. Cho, J. H. Kim, K. T. Chong, Bu-net: Brain tumor segmentation using modified u-net architecture, Electronics 9 (12) (2020) 2203.
9. W. Chen, B. Liu, S. Peng, J. Sun, X. Qiao, S3d-unet: separable 3d u-net for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 358–368.
10. S. Peng, W. Chen, J. Sun, B. Liu, Multi-scale 3d u-nets: An approach to automatic segmentation of brain tumor, International Journal of Imaging Systems and Technology 30 (1) (2020) 5–17.
11. X. Feng, C. Meyer, Patch-based 3d u-net for brain tumor segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2017.
12. Q. Zhang, Z. Cui, X. Niu, S. Geng, Y. Qiao, Image segmentation with pyramid dilated convolution based on resnet and u-net, in: International Conference on Neural Information Processing, Springer, 2017, pp. 364–372.
13. Z. Lei, L. Qi, Y. Wei, Y. Zhou, Infant brain mri segmentation with dilated convolution pyramid downsampling and self-attention, arXiv preprint arXiv:1912.12570 (2019).
14. J. Li, Z. L. Yu, Z. Gu, H. Liu, Y. Li, Mman: Multi-modality aggregation network for brain segmentation from mr images, Neurocomputing 358 (2019) 10–19.
15. E. Tiu, Metrics to evaluate your semantic segmentation model, Towards datascience, recuperado de: https://towardsdatascience. com/metrics-to-evaluateyour-semantic-segmentation-model-6bcb99639aa2#:˜: text= Simply% 20put% 2C% 20the% 20Dice% 20Coefficient, of% 20pixels% 20in% 20both% 20images 8 (2019).