

# Automatic 3D Medical Image Segmentation with Densely-Connected Volumetric ConvNets

Qikui Zhu<sup>1</sup>, Bo Du<sup>1</sup> and Pingkun Yan<sup>2</sup>

<sup>1</sup> School of Computer, Wuhan University, Wuhan 430079, China

<sup>2</sup> Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA

QikuiZhu@163.com

**Abstract.** The Automated medical image segmentation in 3D medical images play an important role in many clinical applications, such as diagnosis of prostatitis, medical image cancer and enlarged medical image. However, it is still a challenging task due to the complex background, lacking of clear boundary and various shape and texture between the slices. In this paper, we propose a novel 3D convolutional neural network with densely-connected layers to automatically segment the medical image. Compared with other methods, our method has three compelling advantages. First, our model can effectively detect the medical image region in a volume-to-volume manner by utilizing the 3D convolution rather than the 2D convolution, which can fully exploit both spatial and region information. Second, the proposed network architecture alleviates the vanishing-gradient problem, strengthens the information propagation between layers, overcomes the problem of overfitting and makes the network deeper by adopting a densely-connected manner. Third, besides the densely-connected manner inside each block, we also adopt the long connections strategy between blocks.

**Keywords:** Medical Image Segmentation.

## 1 Method Described

### 1.1 Data Preprocessing

For image preprocessing, we truncated the image intensity values of all scans to the range of  $[-200, 500]$  HU to remove the irrelevant details. To reduce the overall computation time, a simple Unet architecture is trained to get a quick but coarse segmentation of Kidney segmentation. With the region of interest (ROI), our proposed ConvNets efficiently probes intra-slice and inter-slice features for accurate kidney and tumor segmentation.

## 1.2 Proposed Model

Inspired by the DenseNets, in this paper, we propose a novel network architecture with densely-connected layer, as shown in Fig 1, for automated medical image segmentation.

The proposed model contains two parts: Encoding Part and Decoding Part. The encoding part includes one convolutional block and four dense blocks, which consists of a compression path aims to extract features from the data, reducing the resolution and increasing the receptive field by an appropriate stride. The decoding part contains three dense blocks, one convolutional block and one output block, which consists of an expansive path which up-sampling the features maps to recover its original size.

In the encoding part, the first convolution block is constructed by one Conv(3x3x3)-BN-ReLU and one average-pooling layer with stride 2, and each dense block consists of various number of BN-ReLU-Conv(1x1x1)-BN-ReLU-Conv(3x3x3) with growth rate  $\gamma$ . After each Conv(3x3x3) layer, a drop layer with 0.3 drop rate is added to overcome overfitting problem. To reduce the number of input feature maps, a bottleneck layer is applied before two contiguous dense blocks. The bottleneck layer includes one BN-ReLU-Conv(1x1x1) and one average-pooling layer with stride 2.

In the decoding part, before each block, a 3D-Upsampling operation is used to recover the feature resolution. Besides the connection between contiguous blocks, we also apply long connections between the encoding part and decoding part. The long connections connect the block with same resolution in the encoding part and decoding part. Those connections bring many advantages. First, this can effectively propagate context and gradient information both forward and backward during training. With the network going deeper, the network always encounters the vanishing-gradient problem, the connections between the encoding part and decoding part can directly propagate the gradient information to earlier layers and resolve this problem. Second, this can alleviate the problem of information loss. In the encoding part, when the feature map passes the average pooling layer, part of feature information is abandoned and many detailed information is lost. This always results in inaccurate boundaries in the segmentation results. After adding the long connections between the encoding part and decoding part, the later block receives the information both from earlier block and shallower layers, and the shallower layers can provide that lost feature information. Similar to encoding part, each dense block consists of different number of BN-ReLU-Conv(1x1x1)-BN-ReLU-Conv(3x3x3) with growth rate 0.3. Drop layer and bottleneck layer are employed to overcome overfitting problem and reduce the number of input feature maps. Overall, the detailed structure of the proposed model is shown in the Table 1.

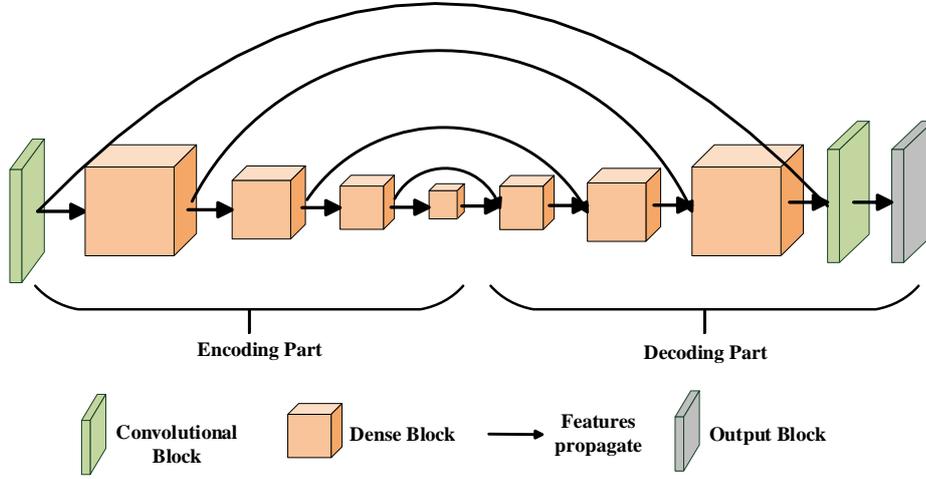


Fig. 1. The architectures of proposed model

Table 1. The detailed structure of the proposed model

	Input size	Output size	Number
Input	128 x 128x16	128 x 128x16	-
Convolution	128 x 128x16	64 x 64x8	-
DenseBlock1	64 x 64x8	32 x 32x4	2
DenseBlock2	32 x 32x4	16x 16x2	4
DenseBlock3	16x 16x2	8x 8x1	6
DenseBlock4	8x 8x1	8x 8x1	10
Up-sampling	8x 8x1	16x 16x2	-
DenseBlock5	16x 16x2	16x 16x2	6
Up-sampling	16x 16x2	32 x 32x4	-
DenseBlock6	32 x 32x4	32 x 32x4	4
Up-sampling	32 x 32x4	64 x 64x8	-
DenseBlock7	64 x 64x8	64 x 64x8	2
Up-sampling	64 x 64x8	128 x 128x16	-
Convolution	128 x 128x16	128 x 128x16	-
Output	128 x 128x16	128 x 128x16	-

### 1.3 Model Training

Before training, we normalized those images as zero mean and unit variance. During training the network, we employ a randomly cropping strategy. We randomly cropped 16x128x128 sub-volumes from the training data every iteration. And to overcome the overfitting problem, we utilize data augmentation to augment the training data. The augmentation operations include rotation, zoom and flip. Our network is trained end-to-end with Stochastic Gradient Descent (SGD) method, and the framework is

implemented under the open-source deep learning library Keras. In the training phase, the learning rate is initially set as 0.0001 and decreased by a weight decay. The momentum is 0.9, and due to the limitation of the memory, we choose 8 as the batch size. Experiments are carried out on GTX1080ti GPU with 11GB memory. For more details, you can read our published paper [1].

## References

- [1] Zhu, Qikai, et al. "A Deep Learning Health Data Analysis Approach: Automatic 3D Prostate MR Segmentation with Densely-Connected Volumetric ConvNets." 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018.