Dice Loss Function for image Segmentation Using Dense Dilation Spatial Pooling Network

Qiuhua Liu janeliu @ruc.edu.cn

Min Fu fm20080525@ruc.edu.cn

January 2018

Abstract

We propose our segmentation model based on Convolution Neural network. Facing the imbalance data problem in segmentation, we adopt DSC loss as loss function when training our network. Apart from exploring its benign properties in theory, we design a novel model to show its effect. Drawing the ideas from dilation convolution [1] and dense network [2], we propose the dense dilation spatial pyramid pooling structure in our network as well as encode and decode network. In order to train out model, we do our experiments in the dataset of PROMISE 12.

1 Method

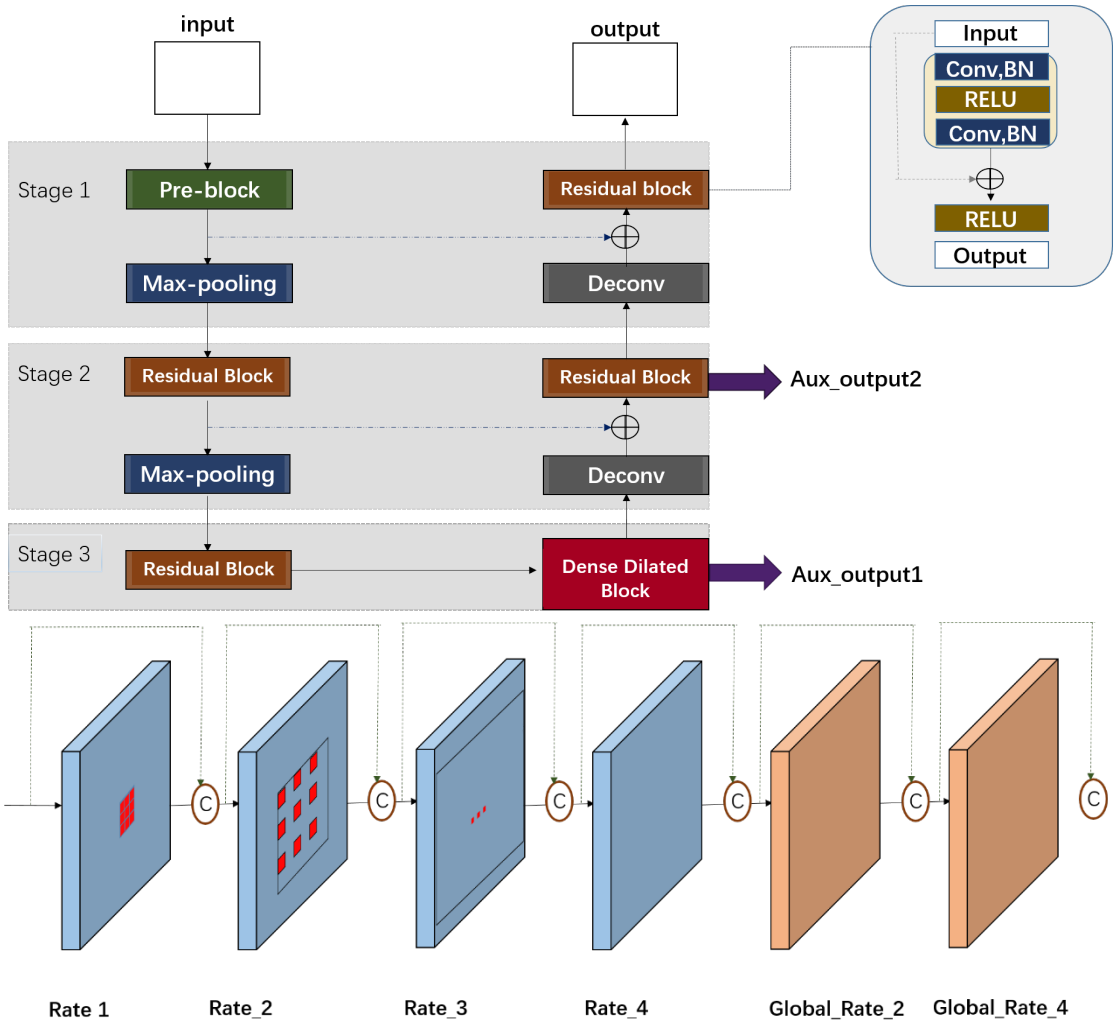
1.1 Pre-process

The training dataset in PROMISE 12 contains 50 volumetric transversal T2-weighted MRIs of prostate, which also includes the ground-truth. For independent evaluation, the testing dataset contains 30 MRIs and their ground-truth are held by the organizer. During the pre-processing step, similar to V-net [3], firstly we use the N4 bias filed correction function of the ANTs framework to normalize the datasets and then resample them to a common resolution of 1*1*1.5 mm. By varying the position of the control points with random quantities obtained from Gaussian distribution with zero mean and 15 voxels standard deviation, we apply random deformations to the training scans.

1.2 Network architecture

Inspire by some outstanding approaches in semantic and medical segmentation, such as Fully Convolution Network and U-Net model, we proposed our RDD segmentation network [figure

one]. The whole RDD network consists of three parts, including a encode subnet, a decode subnet and the dense dilated convolution block [figure two]. We would describe each component of our RDD network next.



Combining together into encode subnet, the convolution layers with more than one stride extract features of the input images, and pooling layers do down-sampling and thus reduce the spatial resolution. Oppositely, the decode network consists of deconvolution layers and convolution layers, and the deconvolution layers do up-sampling and thus improve the resolution of feature maps. When passing through the decode network, they can regain the size as input image. According to the diverse resolutions, we divide these two subnets into several stages, where blocks are residual ones with short residual connection.

To achieve more precise results, unlike the U-net [4] simply using the concatenation, we try to use long residual connections element-wisely, which share the features in encode subnet to the same stage (i.e. with the same solution) of decode subnet.

In semantic segmentation tasks, since the model are required to do the pixel-wise classification, we need to preserve contextual information as much as possible. Therefore we employ the dilate convolution [1], also called atrous convolution.

When dealing with the dilation convolution layers and image level feature, contrast to Atrous Spatial Pyramid Pooling structure (ASPP) [5] concating them parallel, we use densely connection structure , so not only we could concate the multi-level features side by side, but

also reuse those dense features.

1.2 DSC Loss

Basing on the Dice Similarity Coefficient, we adopt a novel loss function, Dice loss.

The definition of DSC Loss: DSC Loss=$1-\frac{2{P_\theta}^T P_r}{\|P_\theta\|_2^2+\|P_r\|_2^2}$

Here, for a brief statement, P_θ and P_r mean two probability vector, and $P_r \in X^N, P_\theta \in X^N$. $\mathrm{X} \in [0,1]$. $\|\cdot\|_2$ denotes the l2 norm.

In order to explain the benign properties of DSC loss, we give the theorem one.

Theorem one. Let P_r be a fixed point over X^N. Let z be a random variable over other space Z. Let g: $\mathrm{Z}\times R^d \to X^N$ be a function, that will be denoted $g_\theta(z)$ with z the first coordinate and θ the second. Let P_θ denotes the outcome of $g_\theta(z)$, and $\mathrm{P}_\theta \in X^N$. Actually, Z is the space of target photos, $\mathrm{P}_r(z)$ is the ground-truth of z, and $\mathrm{P}_\theta(z)$ is the output of the network. Then,

1. If g is continuous in θ, so is DSC($\mathrm{P}_r, \mathrm{P}_\theta$).
2. If g is locally Lipchitz and satisfies regularity assumption 1, then DSC($\mathrm{P}_r, \mathrm{P}_\theta$) is continuous everywhere, and differentiable almost everywhere.

Proof: See Appendix A.

2 Evaluation

2.1 Training protocol

When training the network, we adopt stochastic gradient descent (SGD). Due to the limit of the GPU, we set batch size in one. Besides, in the training, we use an initial learning rate of 0.001, weight decay of 0.005 and momentum of 0.99, and the learning rate reduces 80% in every 10000 iteration. Training time of our model ranges between 8 and 10 hours.

2.2 Results

To show the effect of our loss and network, we design several controlled experiments, and we attain the average DSC value of 86.42% in validation set when using the state-out-of-art network.

Acknowledge

This research was supported by State Key Laboratory of Membrane Biology to Xinqi Gong in Renmin University of China.

Reference

[1] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[J]. 2015.

[2] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[J]. 2016.

[3] Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]// Fourth International Conference on 3d Vision. IEEE Computer Society, 2016:565-571.

[4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241.

[5] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241.