Dilated U-Net based Segmentation of Organs at Risk in Thoracic CT Images

Manoj Satya Kumar Gali[1], Neerja Garg[2], Srikanth Vasamsetti[2]
[1]BML Munjal University, Haryana – 122413, India
[2]CSIR-Central Scientific Instruments Organisation, Chandigarh – 160030, India

ABSTRACT

Cancer is one of the leading cause of death across the globe and projection of radiation towards tumor is the standard treatment of cancer. The first step of irradiation is to delineate tumor from organs near to the tumor. Unlike previous methods of CT segmentation, this paper proposed a procedure for segmentation of organs individually in CT images of the thoracic region such as Heart, Aorta, Trachea and Esophagus and merging them to form multi organ Segmented image. The aim of this method is to avoid coarse output from dilated U-Nets. The overlapping issues has addressed by calculating the mode of eight neighborhood of pixels. The performance of the proposed technique tested on 60 CT scans collected from SegTHOR Challenge. Dice ratio and Hausdorff distance have used in evaluation paradigm.

Index Terms— **Hausdorff distance, dilated U-nets, eight neighborhoods, Coarse output, CT Segmentation.**

1. INTRODUCTION

Cancer is the second leading cause of the death across the globe and projection of radiation is a crucial step for treatment of esophageal and Lung cancer. Radiation dose has to be given with great precision because it is a way of delivering packets of high energy (in general X-rays) to kill cancer and reduce side effects. Success rate of radiation therapy, determined by the cell growth of tumor-effected organs and the organs near to the tumor-effected organs (called as organs at risk) before and after the treatment. Radiation does not kill a cell, it destroys the connection between DNA and cell. This leads to abruption in the process of tumor cell division, called as abortive mitosis and sometimes this can happen to organs near the tumor. Therefore, delineation of organs need to done carefully and accurately. Radiation therapy is a treatment of choice for Lung and Esophageal cancer. The irradiation on organs begins with segmentation of target tumor and the organs near the tumor in computed tomography (CT-scans) images.

In general, experts do segmentation manually by intensity levels and anatomical knowledge e.g. Esophagus is located behind Heart, Trachea is above the Spinal cord, etc. The manual process is costly, time consuming and tedious. This leads to evolution of techniques for automatic segmentation of organs to assist the doctor.

Automatic Segmentation of organs is quite challenging and achieving higher accuracy is very difficult due to several factors such as acquiring volumetric data, low contrast images, variable size of organs from patient to patient, similarity between the shapes of organs and over-fitting towards organs with high intensity or better-structured organs.

Recent trends towards development of deep learning architectures is performing quite well as compared to the traditional methods, especially working with large volumes of data and on variety of data such as audio, video, medical, social, sensor, etc. The development of parallel GPU's, publically labeled datasets, powerful frame works like Tensor flow, and Theano became quite accessible in addition to speeding up the training of deep learning models. Deep learning became a fuel for many computer vision problems such as moving object detection, segmentation, motion tracking [1].

However, several works have addressed for automatic segmentation of organs at risk on CT/MRI scans at different parts of body and using deep learning techniques. In a review paper [2], the techniques of segmentation and detailed algorithm's such as region based, Clustering and classification methods and its applications on MRI and CT scans has been explained. Litjens et al. [3], conducted a survey on deep learning in medical image analysis and described the architectures in convolutional neural network plus explained about its application's in medical image analysis. He et al. [4], worked on segmentation of pelvic organ segmentation using distinctive curve guided Fully Connected Neural Networks (FCN) to segment Rectum, Prostate and Bladder. Segmentation of organs at risk in Thoracic CT images by applying sharp mask techniques with FCN followed by Conditional Random Fields(CRF) is proposed by Trullo et al. [5]. However, the results need to improve furtherly to aid surgeons. Herein, the authors have experimentally showed the success rate of architecture with standard dilated U-net. Atrous convrolution/Dilated Convolutional network [6] model is applied on each organ and using Dice ratio and Hausdorff distance as evaluation metrics. To overcome the challenges mentioned above, this work proposed a procedure for segmentation of organs individually in CT images of the thoracic region i.e. Heart, Aorta, Trachea and Esophagus and merging them to form multi organ segmented image.

2. METHOD

To segment organs (Esophagus, Trachea, Heart and Aorta) accurately from the raw CT images, we are using dilated/ Atrous convolution and the schematic diagram is shown in Fig 1. Proposed architecture is selected based on two reasons 1) the size of organs varies from patient to patient and slice to slice, 2) the output of convolutional neural network in multi segmentation is a coarse output. The detailed explanation of the Dilated U-Net is given in section 2.1. Thereafter, the final segmented output is formed by summation of the individual output coming from each model.

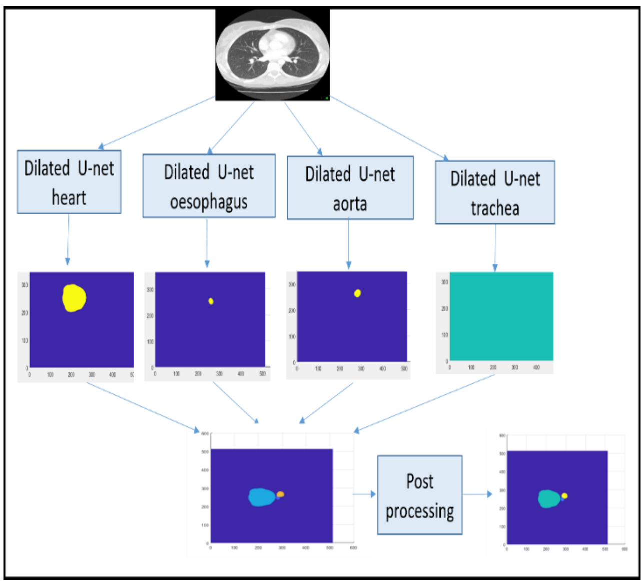

Figure. 1. Architecture for segmentation of organs.

2.1 Dilated U-Net

The use of deep convolutional neural network for fully connection fashioned segmentation has been addressed successfully in [7]. However, the repeated use of average pooling and striding at successive layers reduce the spatial resolution of the output feature maps. The one common approach is to recover spatial resolution in de-convolutional layers as used in [8], but it requires additional time and memory. Papandreou et al. [9], used dilation kernel to generate desirable resolution to feature maps at any layer. It will be applied to a network in two ways 1) post processing technique once network is trained and 2) an integrated model for training. In our case, we have followed the second approach.

Let x[i] be the 1-D input signal, y[i] be the output of dilated convolution of kernel w[k] of length k is

$$y\left[i\right]=\sum_{i=1}^{k}x\left[i+r.k\right]w\left[k\right] \qquad (1)$$

The parameter r corresponds as stride through which we alter the image. Standard convolution is a special case in dilated convolution, if $r = 1$. Dilated convolution helps in enlarging the filed view of kernels at any layer of the Dilated U-Net. Dilated U-Net uses a small kernel (typically 3×3 kernel) in order to control computational time and number of parameters. Dilated convolution with r as rate introduces $r - 1$ zeros in successive values in kernel i.e. enlarging of $k \times k$ kernel into $k_d = k + (k - 1)(r - 1)$ without increasing the computational amount and parameters. It offers a Mechanism to find trade-off between small field of view and large field of view.

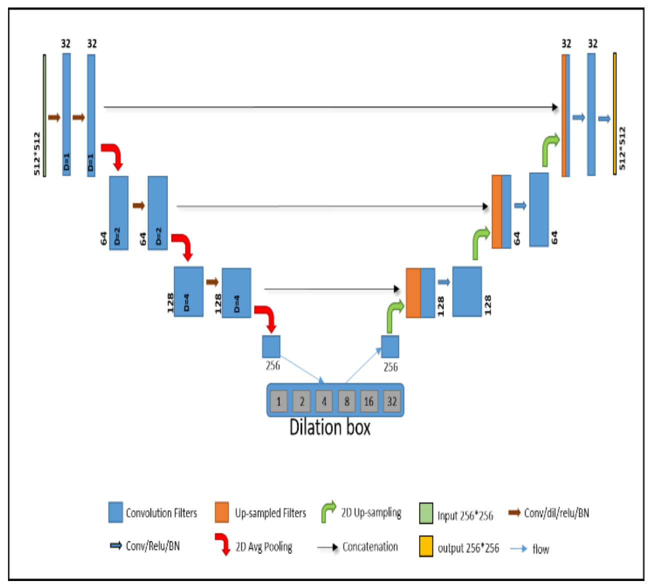

Figure. 2. Dilated U-Net Architecture. The numbers beside the blue boxes is filter size and number represented with D is dilation rate.

We used a Dilated U-Net architecture with 14 layers, in which first six layers involves operation of convolution, dilation, ReLU and batch normalization followed by average pooling after every consecutive two layers. The last six layers are up sampled using bilinear interpolation and concatenation with previous layers. Further, it involves convolution, ReLU and batch normalization. Convolution, ReLU and batch normalization applied on seventh layer and passed to the dilation box individually. The detailed architecture is shown in Figure 2.

The feature maps of seventh layer followed convolution, ReLU and batch normalization passed separately to each layer in dilation box. In dilation box the output of seventh layer convoluted with different dilation rate from $r = 2^0\ to\ 2^5$ and summation of this six layers is given as input to the seventh layer of Dilated U-Net architecture and it is shown in Figure 3.

2.2 Post processing

The outputs from the Heart, Esophagus, Aorta and Trachea are combined to form an overall segmented result. The drawback of this process is overlapping of individual organ regions. To overcome the overlapping regions problem, the mode of eight-neighborhood intensities is used for location $i\ and\ j$ as shown in equation 2

$$\begin{bmatrix} i-1, j+1 & i, j+1 & i+1, j+1 \\ i-1, j & i, j & i+1, j \\ i-1, j-1 & i, j-1 & i+1, j-1 \end{bmatrix} \qquad (2)$$

From this eight neighborhood locations, the mode is calculated and replaces the value at location i, j. By this approach, the overlapping issue is addressed.

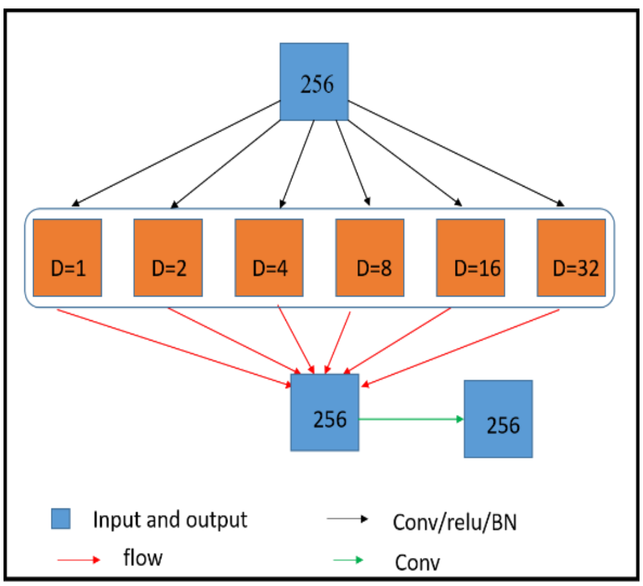

Figure.3. Dilation box working where D represents dilation rate

3. EXPERIMENTAL RESULTS

We performed the experiments on standard dilated U-Net architecture. Further, post processing has applied on Dilated U-Net output using mode as filter. Hausdroff and Dice ratio has used as quantitative evaluation measures. The proposed algorithm has implemented in Python 3.5, 64-bit Windows8 platform with Intel Xenon CPU@2.80 GHz, 64 GB of RAM and 8 GB of GPU.

3.1 Dataset

The performance of the proposed method is evaluated on SegTHOR challenge dataset in CodaLab [5]. We evaluated our model on 60 CT Scans and 40 CT scans for training along with manual segmentation of heart, trachea, aorta and esophagus as ground truth images and 20 CT scans for testing. The size of each scan is $512 \times 512 \times (150{\sim}300)$ voxels and resolution of each scan is $0.98 \times 0.98 \times 2.5$.

3.2 Pre-Processing

Each scan is normalized to zero mean and unit standard deviation. Train test split is performed with test size as 0.2, resulting 32 scans for training and 8 for validation and remaining 20 CT scans for testing.

Data augmentation need to apply when training data is less and it is necessary for network to learn desired properties at microscopic level. In images like CT scans and MRI scans, shift, rotation and deformation invariance is needed and it's implemented in U-Net. We implemented rotation, shift variance-using Keras ImageDataGenerator class focusing on width, height, rotation, and zoom parameters.

3.3 Training

The data is different for each organ i.e. the number of active voxels for each organ is different, so to avoid over-fitting towards a dominant organ we trained each organ separately and summation has done over the intensities of four models. We fine-tune the weights using binary cross entropy loss. Adam gradient descent has used with a learning rate 0.0001 until 75 epochs with 300 epochs per step.

3.4 Results

The results of the proposed method are compared with U-Net architecture and are shown in Figure 4. The U-Net output is shown in Figure 4(b), it is noticed that the output is coarse in nature. Figure 4(c) shows the results of segmentation of Heart, Esophagus and Aorta whereas Figure 4(d) presents the output of overlapping regions separated by applying post processing.

The Table1 gives comparative quantitative evalaution of the proposed U-Net in terms of the dice ratio with U- Net, and proposed U-Net postprocessing. The proposed U-net outperformed basic U-Net in Aorta, Esophagus and Trachea. The proposed U-Net run parllel with Heart. The postproceesing helped to achieve high results in Trachea than proposed U-Net.

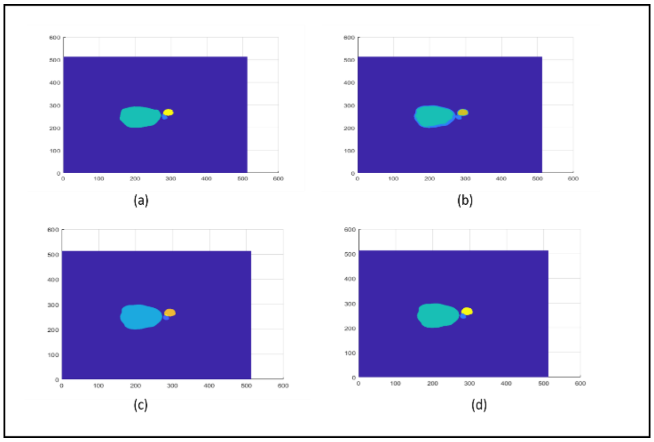

Figure 4. Comparative results of the proposed technique with U-Net method. a) Ground truth b) U-Net Output[10] c) Proposed dilated U-Net d) Output after Post processing

Table 1: Dice ratio for Heart, Aorta, Esophagus, and Trachea

	Heart	Aorta	Esophagus	Trachea
U-Net	0.8562	0.8427	0.3829	0.5536
Proposed U-Net	0.8597	0.8526	0.4648	0.6295
Proposed U-Net Post Processing	0.8595	0.8537	0.4694	0.6425

Table 2: Hausdorff distance for Heart, Aorta, Esophaus and Trachea

	Heart	Aorta	Esophagus	Trachea
U-Net	0.8746	1.3026	3.4941	3.3315
Proposed U-Net	0.8993	1.4577	2.8665	3.9841
Proposed U-Net post processing	0.8930	1.4495	2.8883	2.7342

Table 2 gives the comparative quantitative evalaution of the proposed U-Net in terms of the Hausdorff distance with U- Net, and proposed U-Net postprocessing. The proposed U-net outperformed basic U-Net in Esophagus. The post proceesing helped to achieve high results than basic U-Net and the proposed U-Net in Trachea.

4. Conclusion

In this work, a new framework has been developed for segmentation of Heart, Aorta, Esophagus and Trachea using 60 CT scans dataset from SegTHOR challenge. Individual segmentation of organs from background and augmentation helped to train models on low level and high level features. The results were further improved with a deliberated post processing. For performance evaluation, Dice ratio and Hausdorff Distance metrics were used wherein the segmentation of Esophagus and Trachea shows significant improvement.

REFERENCES

[1]. Voulodimos, Athanasios, N. Doulamis, A. Doulamis, and E. Protopapadakis. "Deep learning for computer vision: a brief review." *Computational intelligence and neuroscience*, 2018.

[2]. Norouzi, Alireza, M.S.M. Rahim, A. Altameem, T. Saba, A.E. Rad, A. Rehman, and M. Uddin. "Medical image segmentation methods, algorithms, and applications." *IETE Technical Review*, vol. 31, no. 3, pp. 199-213, 2014.

[3]. Litjens, Geert, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.V.D. Laak, B.V. Ginneken, and C.I. Sánchez. "A survey on deep learning in medical image analysis." *Medical image analysis*, no.42, pp. 60-88, 2017.

[4]. He, Kelei, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen. "Pelvic organ segmentation using distinctive curve guided fully convolutional networks." *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 585-595, 2019.

[5]. Trullo, Roger, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen. "Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields." *IEEE 14th International Symposium on Biomedical Imaging*, pp. 1003-1006, 2017.

[6]. Chen, L. Chieh, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834-848, 2018.

[7]. Wu, Xundong. "Fully convolutional networks for semantic segmentation." *Computer Science* , 2015.

[8]. Long, Jonathan, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.

[9]. Papandreou, George, I. Kokkinos, and P.A. Savalle. "Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 390-399, 2015.

[10]. Ronneberger, Olaf, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation." In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241. Springer, Cham, 2015.