

ATROUS CONVOLUTION FOR BINARY SEMANTIC SEGMENTATION OF LUNG NODULE

Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, Paul J. Kennedy

Faculty of Engineering and IT
University of Technology Sydney

ABSTRACT

Accurately estimating the size of tumours and reproducing their boundaries from lung CT images provides crucial information for early diagnosis, staging and evaluating patients response to cancer therapy. This paper presents an advanced solution to segment lung nodules from CT images by employing a deep residual network structure with Atrous convolution. The Atrous convolution increases the field of view of the filters and helps to improve classification accuracy. Moreover, in order to address the significant class imbalance issue between the nodule pixels and background non-nodule pixels, a weighted loss function is proposed. We evaluate our proposed solution on the widely adopted benchmark dataset LIDC. A promising result of an average DCS of 81.24% is achieved, outperforming the state of the arts. This demonstrates the effectiveness and importance of applying the Atrous convolution and weighted loss for such problems.

Index Terms— Nodule segmentation, Deep learning, Atrous convolution, weighted loss

1. INTRODUCTION

Lung cancer is recognised as one of the most common types of cancer yet by far the deadliest [1]. Lung nodule analysis and treatment monitoring are effective strategies for early lung cancer diagnosis and survival time improvement. Radiologists usually have to manually look for malignancy in a pool of images and experimentally match cancer related features to candidate tumours. This manual process easily leads to some features being missed in many cases, yet there exist many features which cannot be extracted visually, but can be with an automated system. Inter-observer variation is another major problem of manual nodule segmentation. The disagreement between radiologists in segmenting nodules is reported to be between 2% to 49% [2]. The above mentioned facts are the primary motivation of developing an automatic segmentation model capable of segmenting the nodule accurately with a robust performance.

Recently, Convolutional Neural Networks (CNNs) based deep learning technologies have revolutionised many fields

including semantic image segmentation. Semantic segmentation is a computer vision task to recognize, localize and understand what is in the image at the pixel level. Deep learning based medical image segmentation has been used widely in organ segmentation to segment target organs from the neighbouring tissues [3, 4] as well as in tumour and lesion segmentation for cancer detection and staging [5, 6]. Specifically, in lung tumour segmentation, deep learning based methods have achieved significant improvement, such as in [7, 8]. Since the introduction of the Fully Convolutional Neural Networks (FCNs) [9], the performance of CNNs in semantic segmentation has been improved significantly. One of the most famous derivations of the FCN in medical image segmentation is the U-Net [10].

The heterogeneous appearance of lung nodules and also the presence of similar characteristics between nodules and their surroundings is a serious technical challenge in developing a robust segmentation model. A common solution is to have multiple re-scaled versions of the same image [11, 12] and perform the segmentation on multi-scale images and combine the results. Although it has been noted that this approach can boost accuracy, it comes at the cost of the increased computational load.

Nam et al. in [13] utilized the nodule diameter information to implement a sampling loss. In their approach, instead of calculating the loss for all the pixels of a patch, N random pixels within the patch are selected and evaluated. The information about the nodule diameter helped to limit the radius in which the random pixels could be chosen. In [8], a central

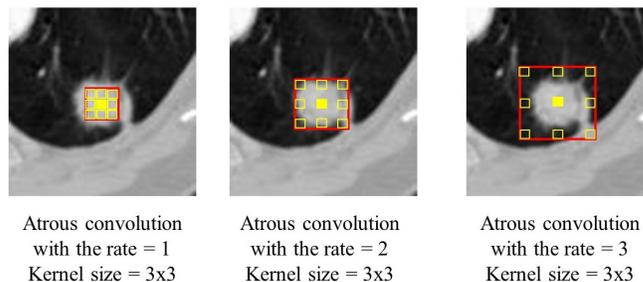


Fig. 1. Illustration of the Atrous convolution with various rates showing enlarged field of view.

M. H. Hesamian, W. Jia and X. He are with the School of Electrical and Data Engineering. P. Kennedy is with the School of Software.

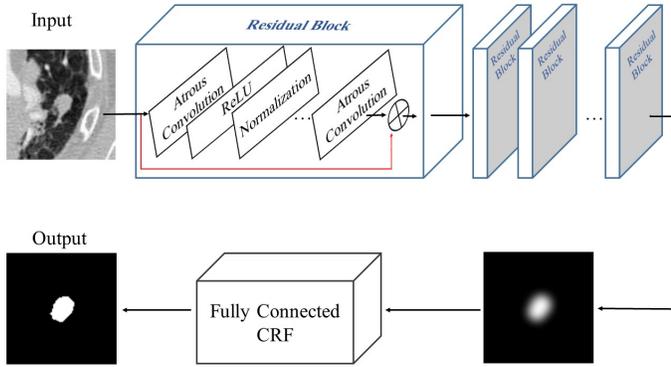


Fig. 2. The block diagram of the proposed network.

pooling layer was proposed, which preserved most features related to the centre of the patch and eliminated the redundant edge features. Moreover, 2D and 3D pathways were developed to separately analyse the 2D and 3D samples, and at the end, corresponding feature maps were combined for final evaluation, which yielded a significant improvement in lung nodule segmentation.

Both of the above mentioned methods, [13] and [8], are still bound to a fixed receptive field size for the kernels which affects the localization power of the network. In our work, we explore the state-of-the-art deep residual structure with Atrous convolution, to increase the field of view of the filters and improve the recall rate. Moreover, through our development, we observed the significant class imbalance issue between the nodule and non-nodule regions, for which we propose a weighted loss function.

The main contributions of the proposed model are in three folds: 1) Introducing the Atrous convolution to increase the field of view of the filters. 2) Applying very deep network structure to improve the segmentation accuracy. 3) Applying a weighted loss to address the class imbalance issue associated with lung nodule segmentation.

The rest of the paper is organised as follow: Section 2 details the methods used in this study and Section 3 analyses the achieved results. The paper concludes in Section 4.

2. THE PROPOSED METHOD

2.1. Atrous Convolution

Scale variation in lung nodules is a challenging problem which has been addressed in other works by simply training the model on a dataset containing varying scales of tumours. The Deep Convolutional Neural Networks (DCNN) usually employ small kernels for convolution to keep the number of parameters and computation in balance. Such small kernels unavoidably cause the loss of localization power.

An Atrous convolution, as illustrated in Fig. 1, with the rate of r will increase the field of view of a kernel of $k \times k$ to $k_e = k + (k - 1)(r - 1)$ without increasing the number of parameters and computation. They can provide more resolution enhancement, and also help to enlarge the field of view of filters [14], which enables the filters to consolidate more contextual information. This property of Atrous convolutions will help to mitigate the object scale issue by extracting score maps from multiple re-scaled versions of original input.

2.2. Weighted Loss

In training a deep neural network for semantic segmentation of medical images, there is typically a significant issue of class imbalance, i.e., the number of pixels belonging to different classes in training samples are significantly imbalanced. Since the target class, i.e., the nodule region, only occupies a small portion of the image and most of the image area is background, the model always gets heavily biased to the background class. In such condition, the model will ignore the weights related to minority class after just a few steps of training and consequently, deliver a very poor output on the target class.

One way to combat this problem is patch-wise training combined with patch selection, as proposed in [8]. In this approach, during the creation of the training set, a control mechanism can be set to have a balanced number of patches from the background and foreground [3]. The same strategy can be used to balance the number of pixels belonging to each class within the patch itself. In this solution, only limited surrounding area (background) around the tumour will be extracted to build up the patch [13, 15].

Another solution for such a class imbalance issue is sampling loss in which N random pixels are selected, and the loss will be calculated just on selected pixels instead of calculating for the entire image pixels [13]. However, the effectiveness highly depends on the value of N , to be a good representative of all pixels.

In our method, we apply a weighted loss. The general idea of weighted loss is to push the network to calculate the loss by assigning different coefficients to tumour and non-tumour pixels, so that the loss for minority tumour and majority non-tumour pixels will not be treated equally. This means by applying a higher ratio to the tumour pixels, their corresponding loss will contribute more to the overall loss and hence, the network is forced to learn more from the foregrounds than the background.

To calculate the loss coefficient, tumour and non-tumour related pixels are counted throughout all the training set samples. Then, this ratio is applied to the individual loss of the pixels. Lets get p_i and p_j two pixels belong to sample S_n . Function $B(x)$ determines the belonging class of pixel x which could be tumour class = c_1 or non-tumour class = c_2 . y_{true} and y_{pred} are the actual and prediction value for

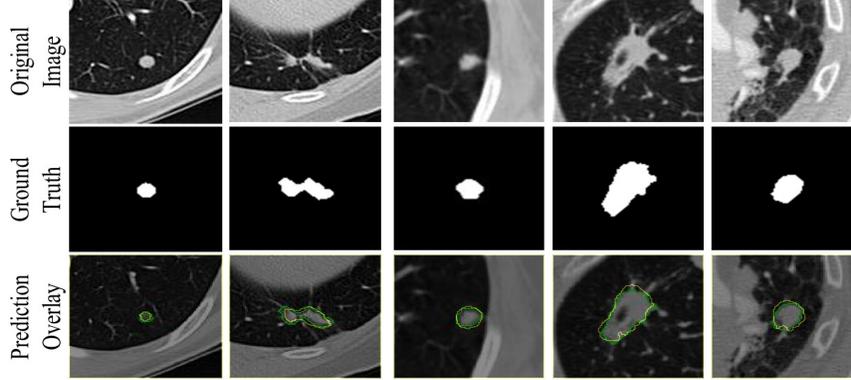


Fig. 3. Examples of the segmentation results, where the green lines indicate the ground truth and the yellow lines indicate the prediction obtained with our approach.

a given pixel. The weighted loss of the sample S_n will be calculated according Eq. 1, where r_1 and r_2 are the ratios of tumour pixels and non-tumour pixels account for throughout all training samples, respectively, and K is the total number of classes, i.e., 2 in this case for two classes.

$$\begin{aligned}
 \text{Loss}(S_n) = & \quad (1) \\
 r_1 \times & \left[\sum_{(p_i \in S_n | B(p_i) = C_1)} \sum_{k=1}^K (-y_{true}^k(p_i) \log(y_{pred}^k(p_i))) \right] + \\
 r_2 \times & \left[\sum_{(p_j \in S_n | B(p_j) = C_2)} \sum_{k=1}^K (-y_{true}^k(p_j) \log(y_{pred}^k(p_j))) \right]
 \end{aligned}$$

2.3. Residual Network

Our proposed structure uses a very deep residual structure for nodule segmentation. Majority of CNN structures are composed of stacking multiple layers each performing a function of F and the output of Y_n , as:

$$Y_n = f(Y_{n-1}; W_n)$$

Theoretically, it is proven that deeper networks have a higher capability to learn [16], as a deep network can represent all functions of a shallow network and some more from the added layers. However, deeper networks suffer from gradient vanishing problem and also the more pressing issue of degradation. He et al. [17] introduced the residual networks, in which instead of consecutively feeding the stacked layers with the feature map, a residual map fed to every few layers. In other words, the residual maps, which are skip connections, allow the network to redirect the derivatives through the network by skipping some layers, as:

$$Y_n = Y_{n-1} + f(Y_{n-1}, W_n)$$

It has been observed that the residual network structure has superior training properties compared to the traditional feed-forward convolutional networks [17].

Resnet101 [17] is used as the deep convolutional neural network for that this structure helps to mitigate the problem of gradient vanishing. In the end, we employ a fully connected conditional random field (CRF) to improve the segmentation accuracy by capturing the fine details. The CRF is jointly trained with the DCNN to build an End-to-End training system (Fig 2). The CRF helps the model to get a sharper segmentation at the boundary pixels.

3. EXPERIMENTS AND DISCUSSION

To demonstrate the feasibility of our proposed solution, we trained and tested our model on the LIDC dataset [18]. In this section, we first briefly introduce the dataset and training procedure, and then quantitative and qualitative experimental results are presented.

3.1. Data

The LIDC dataset [18] contains 1024 cancer (positive) CT cases, each case may have 100 to 400 slices depending on the physical characteristic of patients. The LIDC dataset has the full annotation of the slices separately reported by four different radiologists. The annotation has both the position of the tumour centre and the pixel location of the tumour boundary.

The input to the model are patches of $100 \times 100 \times 3$ pixels, with the greatest in-plan dimension of 3mm and above. All inputs are assumed to contain the tumour, as the focus of this work is the segmentation. To create the patches an imaginary bounding box is drawn around the tumour and then the patch is extracted from the original image.

To create an 8-bit RGB image from the DICOM image, we mapped the Hounsfield Units (varies from -2000 to 4000

Table 1. Performance comparison of different models on their mean Dice score and Precision \pm standard deviation

Methods	Dice (%)	Precision (%)
Graph Cut [8]	68.90 \pm 16.03	65.09 \pm 22.42
U-NET [8]	79.50 \pm 13.95	87.18 \pm 16.13
Simple Diameter [13]	78.78 \pm 18.68	74.17 \pm 23.68
CF-CNN (2D) [8]	80.47 \pm 10.76	75.84 \pm 13.14
Our method	81.24 \pm 1.40	79.75 \pm 4.08

HU) to the range of 0 to 255. Then, we created a standard gray image from the generated matrix.

We perform image normalisation by subtracting each colour channel from its corresponding mean image, which is collected by averaging the whole dataset values. Since the single grey image is duplicated in all three channels, the same mean image is deducted from all three channels. The image normalisation will help to improve the network convergence speed and reduce the training time.

3.2. Training

For training of the network, we have employed the patch-wise training strategy from scratch. Although many researchers believe that the transfer learning may result in better performance, there are some experiments showing that training from scratch delivered better results compared to fine-tuning a pre-trained network [19].

All the experiments were run on Linux (REH7.0) using Nvidia Quadro P4000 GPU with 8GB of memory and the Tensorflow library (version 1.8.0).

3.3. Experiment results

To evaluate the accuracy of the system we measure the Dice score of segmentation, denoted as DSC , which is defined as:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \text{ or } DSC = \frac{2TP}{2TP + FP + FN} \quad (2)$$

In Eq. 2, TP represents the number of the pixels which are correctly classified into the correct classes X and Y , and similarly FP are the number of pixels which are misclassified.

Moreover, the average precision is also measured to enable us to compare our model with the state of the arts, which is defined in Eq. 3:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

We used 300 scans for the training of the network which is only one-third of the samples used in [13] and [8] which are trained on full dataset. The lower number of samples being used demonstrates the efficiency of our model and also reduces the dependency on annotated data. The samples are randomly split to train and validation tests. To validate our

results we performed 10-fold cross-validation. Visualised results of the segmentation by our model allows us to see the output of network against different types of nodules.

In Fig. 3, some example nodules and their corresponding segmentation results are shown to show the performance of the network on different types of nodules.

As the LIDC dataset consists of different types of nodules, our low standard deviation emphasises on the robustness of the system against different types of nodules. Moreover, using a deep network structure not only helps to improve the segmentation performance but also increases the robustness of the network. Thus, as it can be seen in Table 1, the standard deviation of our results is much lower compared to others, which shows improved network stability.

Arguably, the high False Positive rate is the most common issue of lung nodule segmentation models, and it is a key element which discriminates two networks. Our second-ranked precision results represent the lower FP compared to other methods. The proposed model by incorporating more contextual information and also deep residual network was able to reduce the false positive.

By analysing the results pertained in Table 1, it can be observed that we achieved the highest dice score and very competitive precision. In other words, our model tends to predict larger area compared to actual tumour area. This could be considered as the side effect of atrous convolution in tumour segmentation.

4. CONCLUSION

In this study, we have presented a deep CNN model for lung nodule segmentation. We employed the Atrous convolution to increase the receptive field of the filters and processed more contextual information at the same computational cost. We employed a very deep model for the segmentation which can learn more and delivered higher accuracy. Meanwhile, we addressed the class imbalance issue and small organ segmentation in deep network structures by applying the weighted loss. The model was tested on the public benchmark dataset LIDC and has been demonstrated to deliver an average Dice score of 81.24 with very low standard deviation of 1.40.

5. REFERENCES

- [1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal, "Cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [2] Adrian P Brady, "Error and discrepancy in radiology: inevitable or avoidable?," *Insights into imaging*, vol. 8, no. 1, pp. 171–182, 2017.
- [3] Fausto Milletari, Seyed-Ahmad Ahmadi, Christine Kroll, Plate, et al., "Hough-CNN: deep learning for segmentation of deep brain regions in mri and ultrasound," *Computer Vision and Image Understanding*, vol. 164, pp. 92–102, 2017.
- [4] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu, "An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 111–119.
- [5] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 168–172.
- [6] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, et al., "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.
- [7] Wafaa Alakwaa, Mohammad Nassef, and Amr Badr, "Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)," *Lung Cancer*, vol. 8, no. 8, 2017.
- [8] Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, and Jie Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Medical image analysis*, vol. 40, pp. 172–183, 2017.
- [9] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas 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.
- [11] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.
- [12] Konstantinos Kamnitsas, Liang Chen, Christian Ledig, Daniel Rueckert, and Ben Glocker, "Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI," *Ischemic stroke lesion segmentation*, vol. 13, pp. 46, 2015.
- [13] Chang-Mo Nam, Jihang Kim, and Kyong Joon Lee, "Lung nodule segmentation with convolutional neural network trained by simple diameter information," 2018.
- [14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan 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.
- [15] Wenqing Sun, Xia Huang, Tzu-Liang Bill Tseng, and Wei Qian, "Automatic lung nodule graph cuts segmentation with deep learning false positive reduction," in *Medical Imaging 2017: Computer-Aided Diagnosis*. International Society for Optics and Photonics, 2017, vol. 10134, p. 101343M.
- [16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on CVPR*, 2015, pp. 1–9.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [18] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, McNitt-Gray, et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [19] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri, "Deep end2end voxel2voxel prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 17–24.