A PIPELINE FOR LUNG TUMOR DETECTION AND SEGMENTATION FROM CT SCANS USING DILATED CONVOLUTIONAL NEURAL NETWORKS

Shahruk Hossain *Suhail Najeeb* *Asif Shahriyar* *Zaowad R. Abdullah* *M. Ariful Haque*

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

ABSTRACT

Lung cancer is the most prevalent cancer worldwide with about 230,000 new cases every year. Most cases go undiagnosed until it's too late, especially in developing countries and remote areas. Early detection is key to beating cancer. Towards this end, the work presented here proposes an automated pipeline for lung tumor detection and segmentation from 3D lung CT scans from the NSCLC-Radiomics Dataset. It also presents a new dilated hybrid-3D convolutional neural network architecture for tumor segmentation. First, a binary classifier chooses CT scan slices that may contain parts of a tumor. To segment the tumors, the selected slices are passed to the segmentation model which extracts feature maps from each 2D slice using dilated convolutions and then fuses the stacked maps through 3D convolutions - incorporating the 3D structural information present in the CT scan volume into the output. Lastly, the segmentation masks are passed through a post-processing block which cleans them up through morphological operations. The proposed segmentation model outperformed other contemporary models like LungNet and U-Net. The average and median dice coefficient on the test set for the proposed model were 65.7% and 70.39% respectively. The next best model, LungNet had dice scores of 62.67% and 66.78%.

Index Terms— Radiomics, Segmentation, CT Scan, Lung Tumor, Deep Learning

1. INTRODUCTION

Lung Cancer is the leading cause of cancer-related deaths worldwide. Most lung cancers are diagnosed too late during the advanced stages where chances of survival are minimal at best [1]. This is especially true for developing countries and remote areas where doctors and equipment are scarce. However, state-of-the-art computer vision techniques and deep learning have provided an opportunity to build automated early screening systems, giving more people a greater chance of beating cancer. To this end, we explore different neural network model configurations and propose a pipeline for the detection and segmentation of Non-Small Cell Lung Cancer in this paper. We used 3D CT scan data from the NSCLC-Radiomics dataset [2, 3, 4] for our experiments. Our approach consists of a binary classifier for detecting tumors and a hybrid-3D dilated fully convolutional neural network for segmenting potential tumor regions. Our proposed pipeline outperforms established segmentation networks [5, 6] which have been traditionally based on 2D slices.

2. RELATED WORK

There has been a significant amount of research related to radiomics and biomedical image processing utilizing computer vision and data driven algorithms in recent times [3, 7, 8]. With the availability of large volumes of data of medical image data as well as technological improvements, deep learning based approaches have taken the lead in most biomedical image processing applications [9, 6, 5]. Regarding tumor segmentation, there have been works incorporating a deformable model with machine learning to perform the delineation of tumor boundaries [10]. Recurrent Neural Networks (RNN) alongside the more traditional Convolutional Neural Networks (CNN) for segmentation of the pancreas from CT and MRI scans were investigated by [9]. Ronneberger et al. [6] set a new benchmark in the segmentation domain by introducing a encoder-decoder convolutional neural network architecture called U-Net. Their network bested all previous state-of-the-art segmentation techniques based on CNN and pixel-wise classification. More recently Anthimopoulus et al. introduced a fully convolutional neural network using *dilated* convolutions for the semantic segmentation of interstitial lung diseases from thoracic CT scans [5]. Their experimental results showed that their architecture outperformed all other models including U-Net.

3. METHODOLOGY

3.1. Data Preparation

The dataset used in this work consists of samples taken from 300 patients from the NSCLC-Radiomics dataset [2, 3, 4]. A scanned 3D volume of the chest region containing the lungs, heart etc. was available for each patient. For all patient cases,

512×512 slices were provided in the traditional DICOM format. Each CT scan had annotated regions made by an expert radiologist. These 300 were divided randomly into two groups - training (260) and test (40). The total number of CT scan slices (axial) available in each group is shown in Table 1.

Table 1. Number of axial CT scan slices in dataset.

Set	Total Patients	Tumor	Non Tumor
Train	260	4296	26951
Test	40	848	3630

The Houndsfield Unit (HU) values for all slices were clipped to 0-1800 and then mapped to values between 0 and 1. A large portion of each slice was empty space (outside of the patient's body). Analyzing the locations of patient bodies in the training set, we decided on a suitable crop range to exclude a significant portion of the empty space. We found this range to be from 74 to 426 along both the x and y axis, resulting in 352×352 slices. Each slice was then resized to 224×224 using cubic interpolation. The number of training slices was increased 7 fold using a combination of rotations, flips and elastic transformations which have become commonplace among biomedical image augmentation techniques [11, 12, 13]. As shown in Table 1, a large percentage of the training CT scan slices (86%) did not contain any part of the tumors. This class imbalance, if left unchecked, would tend to skew the pipeline towards producing many false negatives. In order to combat this, we sampled the training slices into two subsets; one with all tumor containing slices only (subset A), and the other with tumor containing slices plus 10 random non tumor slices from each patient data (subset B). This selective sampling combined with multi-phase training (described below) helped overcome any issues arising from the class imbalance. A subset C was also created where each entry consisted of a 3D stack containing 9 consecutive slices. The use of this subset is described in the following section. Lastly, 10% of each of the subsets A, B and C were held off as validation datasets.

3.2. Segmentation Model

In this paper, we present a new hybrid-3D dilated fully convolutional neural network based on the LungNet architecture proposed by [5], for the problem of segmenting tumor regions. The network architecture is shown in Fig. 1. The original LungNet was made for 2D slices of the CT Scan. This meant the information present in the 3D structure of the scans were not utilized by the network. To take advantage of this, we added 3D convolutional blocks to our model, as shown in Fig. 2. The model utilizes the original 2D LungNet to produce 2D feature maps from individual 2D slices. These feature maps are then stacked and fused through the 3D convolutions which allow encoding of information along the z axis of the CT scans as well as that along the x-y axes into the output.

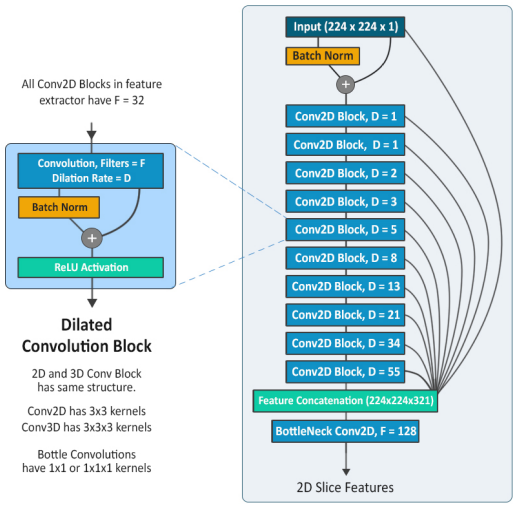

Fig. 1. LungNet feature extraction block.

The model is fed mini-stacks of 9 consecutive CT scan slices. The number of slices in a stack was limited by the computational memory available to us. Each slice is denoted by s_i where i is the slice index corresponding to the depth of the slice in the CT scan. First, 2D feature maps from each of these slices are extracted separately as follows: 10 dilated convolutional blocks with 3×3 kernels and 32 filters each, are applied sequentially on s_i. The dilation rates for each convolution layer are 1,1,2,3,5,8,11,13,34 and 55 along both the x and y axes. The outputs from each block as well as s_i are concatenated in to a single tensor and passed through a bottleneck (1×1 convolution) layer with 128 output filters. We denote the output of the bottleneck as f_i. Thus we have features $\{f_1, f_2, \ldots, f_9\}$ corresponding to input slices $\{s_1, s_2, \ldots, s_9\}$. The f_i are stacked to create a 3D feature volume, which is passed through two 3D non-dilated convolutional blocks, with 64 and 32 filters respectively and $3 \times 3 \times 3$ kernels. These convolutions along all 3 axis incorporate the structural information available along the z-axis from all 9 slices. Finally, one last 3D convolution with a $1 \times 1 \times 1$ kernel is applied to produce 9 single-channel masks $\{m_1, m_2, \ldots, m_9\}$ corresponding to the respective input slices. Each convolution described above except the final layer is followed by a batch normalization layer [14]. The final output layer uses Sigmoid activation while all other layers use ReLU activation [15]. Thus the pixels in the output of the final layer have floating values between 0 and 1, which represent the probability of that pixel being inside of a tumor region.

The proposed model was trained in two phases - I and II. During phase I, only the feature extractor block, that is 2D LungNet, was trained independently to output single 2D

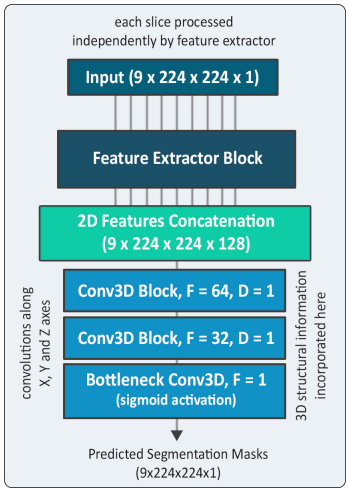

Fig. 2. Proposed model - 3D LungNet.

masks for single 2D input slices. In phase II, the final 3 convolutions were changed to 3D convolutions and the model was trained end-to-end using the mini-stacks to output 9 masks for 9 input slices. Training the full model with random initialization of weights, in a single phase proved difficult in terms of convergence and computational memory. Splitting the training like this allowed the model to converge approximately 4 times faster and improved the overall performance as well. The negative log of the 2D dice coefficient between each ground truth segmentation mask (X) and each predicted mask (Y) was used as the loss function $\mathcal{L}\{X, Y\}$ during training. This loss function proved to be more robust against the class imbalance (tumor region vs. non tumor region) compared to both binary cross-entropy and mean squared error. In phase II, the loss function was averaged over each slice in the mini-stack.

$$\mathcal{L}\{X, Y\} = -log\left(\frac{2 * |X \cup Y| + 1}{|X| + |Y| + 1}\right) \quad (1)$$

We utilized the Adam optimizer [16] with an initial learning rate of 0.001 and batch size of 10 for phase I, while in phase II, learning rate and batch size was set to 0.0001 and 2 respectively. In phase I, training was first done using subset A. When the validation loss function plateaued, the model was adapted for non-tumor slices by training with subset B with a decimated learning rate. Solely being trained on subset A, the model produced a lot of false positives - a non empty segmentation mask even though no tumor was present. Again, if only trained on subset B, the model was biased to produce false negatives - empty segmentation mask even though tumor was present). The aforementioned strategy reduced the bias of the model. For phase II, subset C was used to train model when 3D convolution layers were added.

In the test phase, for a single CT scan, we produced mini-stacks from consecutive overlapping sequences of slices. The output sequences were averaged where they overlapped, thus providing an "ensemble" effect, improving the accuracy of segmentation. The overlapping sequences were formed as follows. If a tumor was detected between slices with index $i = a$ and $i = b$ where $a < b$, the first mini-stack consisted of s_a, 4 slices above and 4 slices below s_a respectively arranged in order of slice index. That is, the mini-stack contained $\{s_{a-4}, s_{a-3}, \ldots, s_a, \ldots, s_{a+3}, s_{a+4}\}$. The next mini-stack had all indexes shifted up by 1. This was done until the last slice in a mini-stack was s_{b+4}. The mini-stacks were padded with the nearest slice if we hit the top/bottom slice of the CT scan.

3.3. Binary Classifier

A vast majority of slices in the training set did not contain any part of the tumor. We therefore implemented a frontend binary classifier, shown in Fig. 3, to detect tumor containing slices and only pass those to the segmentation model. Similar to the segmentation model, features were extracted using the already trained feature extraction block and then a binary decision was made through fully connected dense layers. To boost the recall of the binary classifier, when a slice was predicted to contain a tumor, 8 neighboring slices along the $\pm$z directions were also automatically labeled as having parts of the tumor. Each slice represented a real world thickness of 3 mm as per the settings of CT Scanner. So, labelling 8 neighboring slices as tumor containing meant labelling a region spanning of $2 \times 8 \times 3$ mm = 48 mm along the z axis to be tumor containing. This made all but the largest tumors fall into the range of slices we would be looking to segment. Following this scheme the binary classifier had only 40 false negatives and 1158 false positives for the test set. Most of the false positives, however, were weeded out by the segmentation model and post processing.

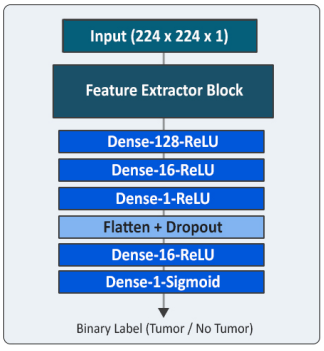

Fig. 3. Tumor detection model.

3.4. Post Processing

The masks generated by the segmentation model were padded (to compensate for the cropped regions) and resized back to the original resolution of 512×512. An optimum threshold

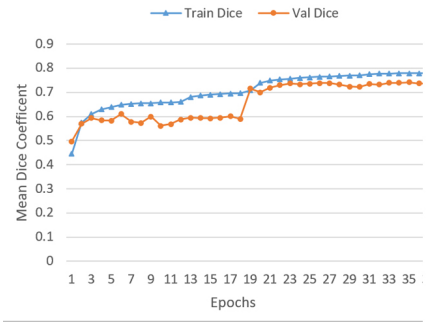

Fig. 4. Training curve for feature extractor.

value was determined from the Receiver Operating Characteristic (ROC) curve generated from the validation data, and applied on the masks - values below it were relegated to 0, while those above were made 1. The value for the proposed model was found to be 0.75. To remove noisy specks and small patches, we applied a sequence of dilation and erosion operations with elliptical and cross shaping kernels. We also applied an area-based threshold for the predicted tumor regions. Regions smaller than the threshold (25 px^2 was found to be the smallest tumor region in the training set), were relabeled to be non-tumor. These post processing steps reduced false positives by more than 50%. The post-processing steps were applied to the output of all the tested segmentation models to enable a fair comparison of the performance of the segmentation models.

4. EXPERIMENTAL SETUP AND RESULTS

4.1. Training and Evaluation

We implemented our models using the Tensorflow framework by Google [17] running behind the wrapper library Keras [18]. Pre and post processing algorithms were implemented using functions from OpenCV and Scipy libraries for python [19, 20]. By following the training scheme described in section 3.2, we obtained the training curve shown in Fig. 4 for the Segmentation model in phase I. The sudden rise in dice coefficient of the model corresponds to the time when training on subset B started. It is seen that the model generalizes well and does not overfit. The final validation dice coefficient averaged over each patient case was 62.7%. For phase II, the model was trained on subset C for 6 epochs at which it reached peak validation dice coefficient.

4.2. Results

The proposed segmentation model was compared against the 2D LungNet and U-Net architecture based on the 2D dice coefficient. The dice coefficient was calculated for each tumor containing slice and then averaged for all cases in the test set. These results are presented in Table 2. The average values of the dice coefficient were calculated for only slices having

Table 2. Results summary on test set.

Model Arch.	Total Params. ($\times 10^5$)	Mean Dice (%)	Median Dice (%)
U-Net	310	58.48	62.29
LungNet	1.30	62.67	66.78
Proposed	**4.03**	**65.77**	**70.39**

tumors as indicated by the ground truth annotation. That is, false positives were ignored in the average. It is seen that the proposed model obtained a much higher dice coefficient. The high median dice score also indicates that the proposed model performed better on edge cases, where the other models failed. Some of the generated masks from the test set are shown in Fig. 5. It is seen that visually the proposed model produces a less spiky, more encompassing mask of the tumor region, owing to the incorporation of the 3D structural features.

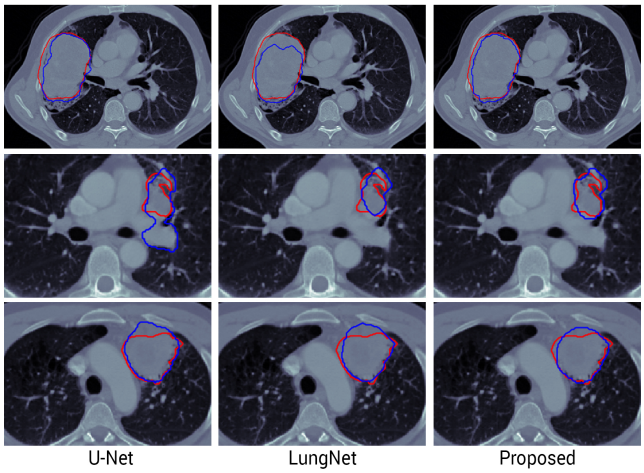

Fig. 5. Segmentation boundaries produced by the pipeline using different models. The blue outline represents the predicted mask while the red outline indicates the ground truth.

5. CONCLUSIONS

The proposed segmentation model was seen to perform better by almost 3% in terms of dice coefficient compared to the next best implemented model. It was able to take advantage of the 3D information present within CT scan volumes. We also emphasize on the fact that the hybrid-training of the feature pooling LungNets ensures an effective yet less computationally expensive architecture compared to end-to-end 3D convolutional model based pipelines. As reported in [21] an earlier version of the proposed pipeline outperformed all other such approaches in the IEEE VIP CUP 2018 Lung Tumor Segmentation Challenge. Further works planned comprise of training the pipeline with deeper stacks of slices and training the binary classifier and segmentation model jointly.

6. REFERENCES

[1] David E Midthun, "Early detection of lung cancer," *Medical Physics*, vol. 5, pp. F1000 Faculty Rev–739, 2016.

[2] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, and Philippe Lambin, "Data from nsclc-radiomics. the cancer imaging archive," 2015.

[3] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, pp. 4006, 2014.

[4] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al., "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[5] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Thomas Geiser, Andreas Christe, and Stavroula G. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *CoRR*, vol. abs/1803.06167, 2018.

[6] 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.

[7] "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[8] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, et al., "Radiomics: the process and the challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1234 – 1248, 2012.

[9] Jinzheng Cai, Le Lu, Fuyong Xing, and Lin Yang, "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning," *CoRR*, vol. abs/1803.11303, 2018.

[10] Eugene Vorontsov, Nadine Abi-Jaoudeh, and Samuel Kadoury, "Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Springer, 2014, pp. 74–83.

[11] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, 2003, pp. 958–963.

[12] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2018, pp. 230–234.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, pp. 234–241, Springer International Publishing.

[14] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.

[15] Vinod Nair and Geoffrey E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.

[16] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[17] Martin Abadi, Ashish Agarwal, Paul Barham, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, Software available from tensorflow.org.

[18] Francois Chollet et al., "Keras," 2015.

[19] G. Bradski, "The opencv library," *Dr. Dobb's Journal of Software Tools*, 2000.

[20] Eric Jones, Travis Oliphant, Pearu Peterson, et al., "Scipy: Open source scientific tools for python," 2001.

[21] A. Mohammadi, P. Afshar, A. Asif, K. Farahani, J. Kirby, A. Oikonomou, and K. N. Plataniotis, "Lung cancer radiomics: Highlights from the IEEE Video and Image Processing Cup 2018 Student Competition SP Competitions," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 164–173, Jan 2019.