Residual-based 3D Faster RCNN For Lung Nodule Detection

Minchao Li Yiheng Zhang Yiqing Wang
Kc-4i group 13

***Abstract*—Detection lung nodule has always been a challenge topic in medical diagnosis and requires a lot of manual work. Recently, the outgrowth of convolutional neural network (CNN) draws a lot of attention for its strong and robust ability to address 2D pattern recognition problem. While 3D CNN has been proposed, challenges remain since 3D input consumes rather much memory space and computational resources, making it difficult to be deployed in many situation. We proposed a residual block based 3D Faster RCNN for lung nodule detection, which is both easier to train than traditional 3D Faster RCNN and can provide decent prediction as well.**

***Index Terms*—Pattern Recognition, Medical Diagnosis, 3D CNN.**

I. Introduction

Lung cancer has been a leading cause of cancer-related death in human being. Early stage detection of lung cancer is one of the most important technologies that improve the patients survival possibility. In the field of radiation medicine, a visible small massive tissue in lung, i.e. a lung nodule, is a potential symptom of lung cancer. Traditionally, lung nodule detection and classification are finished by experienced radiologists via artificially analysis on CT scans. With the recent advance in computer-aided-diagnose and popularization of low-dose lung CT scanning, many image-processing based or data-driven automatic methods are developed for the time-consuming task of lung nodule detection and classification.

In this work, we focuses on design a robust automatic diagnose system whose input is a patient's lung CT scan and the output is the annotation of lung nodules including the location and prediction confidence. As this is a diagnose system, our goal is to minimizing the false negative rate without incurring false positive too much. It's naturally to think about data-driven method which learn to annotation a CT scan like an trained radiologist rather than pure image processing method. In fact, most existing robust systems are divided into two stages to leverage both techniques. The system firstly generates the possible module regions and then is taught to reduce the false positive.

We use LUNA16[1] dataset which is a common public lung CT scan dataset contains 888 low-does lung CTs with location and classification annotation for nodules. As 3D nature of CT data, we build U-net for segmentation and 3D CNN for classification and achieved effective accuracy result. Further work is needed to combine the two stage together for a automatic pipeline.

II. Related Work

Nodule detection Early nodule detection is highly dependent on expertise designed features[2] and techniques such as morphological computation, pixel threshold. Recently, some powerful deep convolutional neural networks shows state-of-art performance on data-driven image processing tasks. Fully convolutional networks[3] for semantic segmentation can be used to generate candidate region for the next classification stage. A more natural representation for neural network is bounded box which can be generated from faster R-CNN [4] for object detection and UNet[5]. Due to the 3D nature of CT scans, some work proposed 3D convolutional networks to handle this task. Therefore, for nodule detection, there are pipeline in the form of faster R-CNN followed by 3D CNN classifier[6] [7], 3D U-net[8] and V-net[9].

Nodule classification Early non-deeplearning methods are mainly based on human-designed 2D or 3D feature such as shape and texture[10] [11]. Recently, some deep learning technique is used in nodule diagnosis. Multi-scale CNN[12] is used for nodule-level classification while transfer learning is used for patient-level CT classification[13]. [14] shows that 3D structured CNN perform better on 3D CT data than 2D CNN.

III. Method

In this section, we will discuss the details of our method, including the preprocessing part, the segmentation part and the classification part. Most of the preprocessing is done by using traditional image filter with well-designed parameters. Section III-A will illustrates details of our proposed filter. As for the lung nodule detection, our structure combines the segmentation part and the classification part together by using modified Faster RCNN model (Fast RCNN + RPN)[15], [16], of which details will be provided in Section III-B

A. Preprocessing

The original dataset from Tianchi competition consists of sliced images I of 3D scanned lung model. Value of pixels in I varies in a wide range, making it difficult to segment the lung nodule with raw images as inputs. Thus, preprocessing is necessary for input regularization, which can significantly boost the robustness of the network.

We first have to extract a lung mask to exclude other tissues out of the lung itself, since many of which may have very similar shapes with nodules. Gaussian filter G is adapted to extract the mask of the lung in frequency domain since it

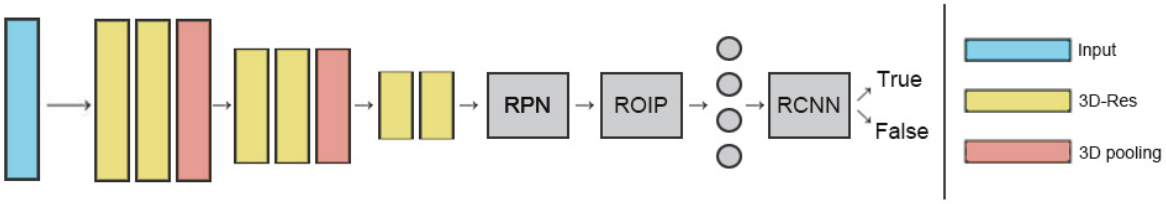

Fig. 1: Overview of our proposed residual block based 3D faster RCNN network structure. We use six residual block (yellow) and two max pooling layer to tranform input(blue) into feature map for RPN block

is comparatively most convenient approach. The definition of spatial Gaussian kernel G with size $2k+1 \times 2k+1$ is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-k-1)^2+(y-k-1)^2}{2\sigma^2}} \tag{1}$$

where σ is the deviation of the 2D Gaussian distribution, which controls the smoothness of the output image.

In our case, we set the kernel size k to 3 and the standard deviation σ to 1. The Gaussian filter produces descent yet blurred result, which requires further processing. We choose to binarize the filtered image with binarization filter B. The definition of a ideal spatial binarization filter B is:

$$B(x,y) = \begin{cases} 0 & I_{x,y} < \sigma \\ 1 & I_{x,y} \geq \sigma \end{cases} \tag{2}$$

where σ denotes the threshold. Here, we choose threshold σ to be -600 as Liao et. al.[16] does for LUNA dataset.

Sadly, there is no direct way to fetch perfectly segmented lung slice for each input image. Thus, for some failure cases, we have to check them manually and design specific parameters to get good filtered results.

After filtering, 2D connected components smaller than $30mm^2$ or with eccentricity larger than 0.99 are discarded to further reduce noise. Then we calculated all 3D connected components. 3D connected components with volume larger than 8.0L or smaller than 0.65L are discarded since they are highly unlikely to be lungs. For failure cases in this part, we again check them manually and deleted some useless samples like void images (images consist of only zero-value pixels). In total, we have 1616 images for training, validation and testing after preprocessing.

B. 3D Residual Faster RCNN

Faster RCNN is an end-to-end object detection network based on Fast RCNN[15], [17]. Unlike Fast RCNN, Faster RCNN uses Region Proposal Network(RPN) to generate rectangle object bounding box instead of using Spatial Pyramid Pooling(SPP). We adopt original 2D Faster RCNN to 3D lung nodule detection with 3D convolution neural network.

We first introduce the definition of convolution layer, pooling layer and fully connected layer in 3D space respectively. 3D convolution layers work on fixed-size $I \times J \times K \times C$ inputs, where the first three dimensions represents the width, height, depth of voxel V. In our case, shape of the preprocessed voxel data is determined, $I = J = 512$, $K = 281$. The fourth dimension of the input indicates number of feature maps. The shape of the convolution filter $C(k,k,k,c)$ is pre determined as well, which creates c feature maps by convolving the input with learned filters of shape $k \times k \times k \times c$. In our case, we use $k = 3$ in each convolutional layer. 3D convolution is also compatible with spatial stride s, which is common in 2D convolution. 3D pooling layers downsample the input with either max pooling or average pooling. We use max pooling in our case. Max pooling downsamples the voxel along the spatial dimensions by replacing each $m \times m \times m$ non-overlapping block of voxels with their maximum respectively. 3D fully connected layer outputs a $n \times 1$ vector which stands for n output neurons. The value of each element in the vector is a learned combination of all the outputs from the previous layer, processed by a nonlinear activation function. We use ReLUs as activation function.

Then it is natural to extend residual block to 3D space as well. The key contribution of residual block and residual learning is to prevent the degradation effect of very deep neural network. Instead of training the neural network to learn to predict direct output, residual block aims to predict residual value $F(x)$ defined as:

$$\hat{y} = F(x) + x \tag{3}$$

where $\bar{y}$ is target result. Residual block performs well in a wide range of classification problem and is efficient to train. Thus we replace traditional VGG structure in faster RCNN with a series of residual block, which not only make it capable for us to build deeper neural network but also boost the convergence speed of RPN network in experiment. In our network, each residual block contains three 3D convolution layer, followed by ReLU activation respectively.

Delighted by RPN block used in 2D faster RCNN[15], we adopt RPN block in our 3D frame work to generate bounding box for lung nodule in 3D voxels. The bounding box in RPN block is represented as:

$$t = (x, y, z, dx, dy, dz) \tag{4}$$

where x, y, z, dx, dy, dz denotes the coordinates and half length of the bounding box's width and height. Each proposed bounding box also has a predict binary label to denotes whether volume in the bounding box contains an object, which is determined by largest IoU of proposed and ground truth bounding box. RPN is trained with position loss, which is back

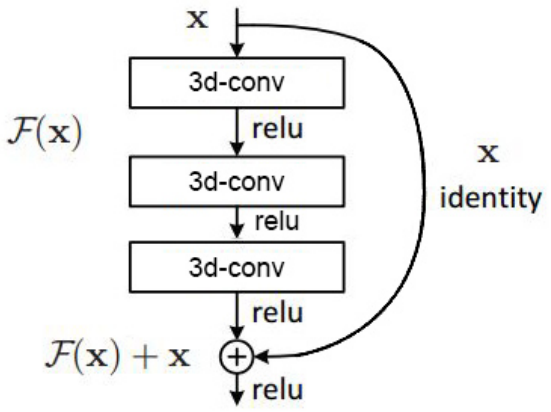

Fig. 2: An illustration of residual block in our model. Three 3D convolution layers are included in one residual block, which are followed by ReLU activation.

propagated by gradient calculated in position loss function $\mathcal{L}_{pos}$. Usually, $\mathcal{L}_{pos}$ is the $\mathcal{L}_1$ loss of t and $\bar{t}$. Here $\bar{t}$ is the bounding box ground truth. Next, the ROIP block transform voxel volume in the proposed bounding box output from RPN to feature space representation by 3D convolution and pooling layer. The last block of our structure is a RCNN block, function of which is a common binary classifier, which has been proved to be good enough in experiments .

C. Loss Design

Last but not least, we will discuss the loss design of our network. Aside from the $\mathcal{L}_{pos}$ loss to tune the RPN block, we need to design classification loss for RCNN classifier. According to our experiment, the false positive samples in proposed bounding box consists of a large part of our dataset (approximately 70%). Usually decreasing the false positive ratio is a prior task in medical diagnosis, thus we have to punish the classification network when it labels false positive to be true. The classification loss $\mathcal{L}_{cls}$ is defined as:

$$\mathcal{L}_{cls} = (1-\lambda)\frac{1}{N_{pos}}\sum_i \mathcal{L}_{pos}(p_i, \hat{p}_i) + \lambda\frac{1}{N_{neg}}\sum_i \mathcal{L}_{neg}(p_i, \hat{p}_i) \tag{5}$$

where N denotes the total sample number, and $\mathcal{L}_{pos}$ and $\mathcal{L}_{neg}$ are $\mathcal{L}_1$ loss of true negative and false positive predictions respectively. λ is a hypeparameter to control the degree of punishment to false positive. In out case, we set λ to 0.7.

IV. Implementation

The experiment was conducted on Ubuntu 16.04.3 LTS with 4 processors, Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.3GHz and 64GB total memory space. Our model is trained on Tesla K80 with 12GB Memory. Except for our 3D residual faster RCNN structure, we implemented U-net as baseline of the segmentation part The Unet framework of uses python 2.7, Kares as the frontend with tensorflow 1.4.0rc0 with Cuda 9.0 as the backend.

As for 3D residual faster RCNN, we choose to use Intel Extended Caffe since the convolution layer, max pooling layer and fully connection layer in caffe is self-adaptive to the shape of input. Some other common libraries used include numpy 1.13.1, SimpleITK 1.1.0, pandas 0.19.2, sklearn 0.18.2.

The training process of U-net is done in 3 hours on Nvidia Tesla K80 graphic card.

V. Experimental Results

The experiment was conducted with LUNA16 dataset. The dataset includes 888 CT scans in total. CT images are stored in MetaImage (mhd/raw) format. Each .mhd file is stored with a separate .raw binary file for the pixeldata. Additionally, 2 csv files are included in the dataset. The file, annotations.csv, contains one finding per line. Each line holds the SeriesInstanceUID of the scan, the x, y, and z position of each finding in world coordinates; and the corresponding diameter in mm. The annotation file contains 1186 nodules. Meanwhile, the file, candidates.csv, contains nodule candidate per line. Each line holds the scan name, the x, y, and z position of each candidate in world coordinates, and the corresponding class (1 represents true positive and 0 represents false positive). For more details about the dataset, please view https://luna16.grand-challenge.org/

Within the preprocessing parts [12], skimage.morphology is implemented to extract the lung area. As different organ tissues react differently to the radioactivation, the HU value in the CT images can be used to extract and eliminate the tissue we want or not. (HU value 604 corresponds to lung [18]) Still, other morphology implementation is used like closing and erosion. Selems used in the function is decided through experiments. The result of preprocessing is shown in Fig. 3

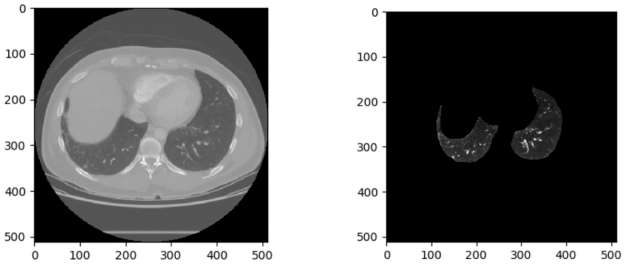

Fig. 3: Comparison between original data and processed data

To implement 3dcnn, vowel should be extracted from the preprocessed ct images. The position (center) of vowel is indicated by candidates.csv. The size of vowel is finally decided as [26,40,40], since with smaller vowel size, too many vowels are invalid bacause they do not contain any tissues (a cubic with all 0s), while with bigger vowel size, the time consumed in training and the pressure given to the hardware is considerable.

There exist 2 methods of extracting vowel, without and with nodule mask. Nodule masks can be generated from annotations.csv. With the first implementation(without nodule mask), the quantity of the extracted vowel is massive and what's worst is that the extracted vowel contains other tissue,

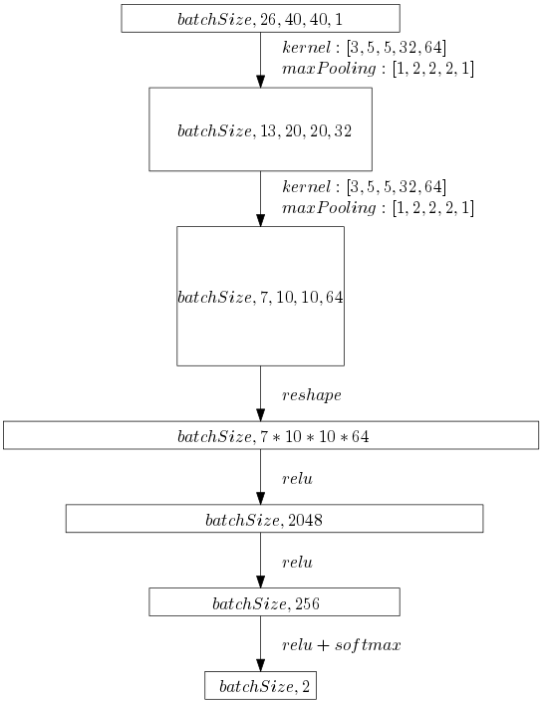

Fig. 4: 3dcnn network structure

like small vessels as is shown in Fig. 5. After visualizing the preprocessed lung structure, it is apparent that without the help of the nodule mask, hardly can we extract the pure nodule in lung (see Fig. 6).

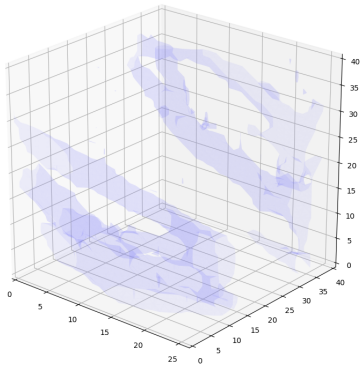

Fig. 5: vowel extracted without nodule mask

Meanwhile, with nodule masks, the lung nodule and vowel extracted performs better, which contains only the nodule (see Fig. 7). However, to implement this method, model used to generate potential nodule mask for new CT image is needed. We implemented U-net to generate the nodule mask. The evaluation of our model on training set is 0.60, that on validation set is 0.58 and that on test set is 0.266. The result is fair enough compared with evaluation value 0.3 given in the

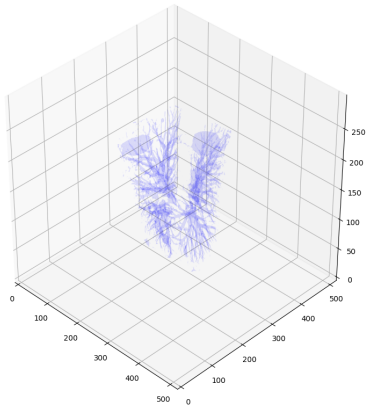

Fig. 6: lung after preprocessing without nodule mask

tutorial. The result shows in Fig.8

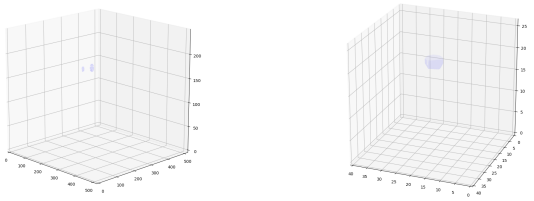

Fig. 7: The left image is the visualization of the entire lung with nodule mask. The right image is the one of the vowels extracted

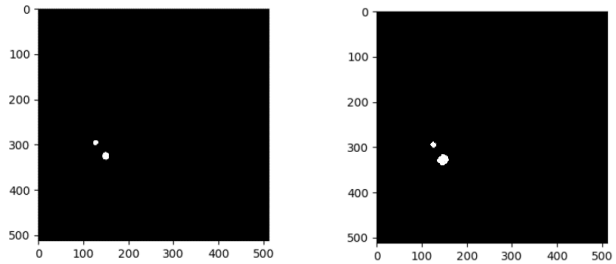

Fig. 8: The left image is original nodule mask. The right image is predicted one

The final step is to train our 3dcnn model with vowels as input and a probability of being true positive as output. Consider the paucity of positive samples, data augmentation is implemented. With every vowel with positive class, by mirroring, shifting and transpose, we can conduct more vowels from one. The structure of our model is shown in Fig. 4, and our final accuracy one the test set is 0.8125.

VI. Conclusion and Future Work

We proposed 3D version of faster RCNN to solve lung nodule segmentation and classification problem. As a multi-task network, faster RCNN provides an end-to-end solution to medical treatment in simultaneous segmentation and classification cases.

For comparison, we also implemented and tested a traditional approach, U-NET for segmentation and 3D CNN for classification. Both of these two methods achieved a descent accuracy in our test set. Considering model efficiency, we affirm that our 3D faster RCNN model is a better choice.

There certainly exists space for conducting future research. First, our research is strongly restricted by computational resources. We only have one PC equipped with E5-2686 v4 CPU and Tesla K80 GPU devices. For researchers interested in lung nodule analysis with comparative abundant computational resources, other very deep structures, cascade structures and multi-network structures can be choices for better test scores. Second, we didn't focus on the tuning of hyperparameters thus we didn't use optimal settings for our network. For some hard cases, hard mining can further increase the robustness of our network.

Acknowledgment

We would like to say thank you for the aid of Professor Sheng, our supervisor Doctor Jason and our dearest TA Doctor Anum Msd. It is their assistance and guidance that we could make these achievements with little background in medical image analysis and 3D deep learning.

References

[1] A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Medical image analysis*, vol. 42, pp. 1–13, 2017.

[2] E. Lopez Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. Fantacci, and P. Cerello, "Large scale validation of the m5l lung cad on heterogeneous ct datasets," *Medical physics*, vol. 42, no. 4, pp. 1477–1489, 2015.

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

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[5] O. Ronneberger, P. Fischer, and T. 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.

[6] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 559–567.

[7] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network," *arXiv preprint arXiv:1711.08324*, 2017.

[8] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.

[10] T. W. Way, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on ct scans: Segmentation and classification using 3d active contours," *Medical physics*, vol. 33, no. 7Part1, pp. 2323–2337, 2006.

[11] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimelfarb, R. Falk, and M. A. El-Ghar, "3d shape analysis for early diagnosis of malignant lung nodules," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 772–783.

[12] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.

[13] W. Shen, M. Zhou, F. Yang, D. Dong, C. Yang, Y. Zang, and J. Tian, "Learning from experts: developing transferable deep features for patient-level lung cancer prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 124–131.

[14] X. Yan, J. Pang, H. Qi, Y. Zhu, C. Bai, X. Geng, M. Liu, D. Terzopoulos, and X. Ding, "Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 91–101.

[15] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[16] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3d deep leaky noisy-or network," *CoRR*, vol. abs/1711.08324, 2017. [Online]. Available: http://arxiv.org/abs/1711.08324

[17] R. B. Girshick, "Fast R-CNN," *CoRR*, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083

[18] B. Sasidhar, "Automated segmentation of lung regions using morphological operators in ct scan," *International Journal of Scientific and Engineering Research*, vol. 4, 2013.