

# BMEN 4000 Final Project Report

Nanyan "Rosalie" Zhu  
*Biotechnology*  
*Columbia University*  
nz2305@columbia.edu

Chen "Raphael" Liu  
*Electrical Engineering*  
*Columbia University*  
cl3760@columbia.edu

**Abstract**—We developed a fully automated computer aided diagnosis (CAD) framework to facilitate the detection of pulmonary nodules in lung computed tomography (CT) scans. The framework consists of a lung-segmentation module that implements a traditional image processing algorithm introduced in [1] and a subsequent deep-learning-based nodule-segmentation module that utilizes the Attention U-Net architecture proposed by [2].

**Index Terms**—Image segmentation, object recognition, object segmentation.

## I. INTRODUCTION

By far, lung cancer has been the leading cause of cancer death each year for both men and women in the US, contributing to a higher annual death count than colon, breast, and prostate cancers combined. In the year 2019, there will be 228,150 new lung cancer cases and 142,670 deaths from lung cancer, according to the estimation given by the American Cancer Society [3]. To reduce the mortality rate of lung cancer, a solution proposed and demonstrated effective in recent years is computed tomography (CT) screening [11]. By interpreting the images generated from thoracic (aka., chest) CT scans, radiologists may catch some early signs of lung cancer, and eventually increase the patients' chance of survival by providing follow-up examinations and treatment plans [7] [8] [9] [10].

Lung nodules, or more formally referred to as "pulmonary nodules", are defined as "round opacity, at least moderately well marginated and no greater than 3cm in maximum diameter" [13]. They are an important feature to which radiologists pay special attention when interpreting the thoracic CT images, as their quantity, size, location and morphology may greatly affect the lung cancer diagnosis. In current clinical practices, human experts are the only entities to view and draw inferences from the CT scan images, and this implies a time-consuming and laborious process of carefully examining beyond 100 slices of images per patient [12].

In the aim to assist the radiologists and alleviate human labor, we developed a computer-aided diagnosis (CAD) frame-

Instructor: Professor Andrew Laine, Biomedical Engineering, Columbia University. (<https://engineering.columbia.edu/faculty/andrew-laine>)

Instructor: Professor Paul Sajda, Biomedical Engineering, Columbia University. (<https://engineering.columbia.edu/faculty/paul-sajda>)

Advisor: Professor Jia Guo, Neurobiology, Columbia University. (<https://jg3400.wixsite.com/sail/post/lab-members>)

Advisor: Professor Christine P. Hendon, Electrical Engineering, Columbia University. (<https://www.ee.columbia.edu/christine-hendon>).

work to automate the process. The goal of the framework is to process any qualified thoracic CT scan and provide a visualization of the scan with the volumes corresponding to the lungs and lung nodules stained in distinct colors. Textual and quantitative analytical results including nodule quantity, volume and morphology may also be provided.

## II. RELEVANT WORK

The need for lung nodule segmentation has received unprecedented attention recently, and there has been many relevant work that address issues similar to what we propose. Several attempts have been made using traditional image processing approaches. For instance, Badura et al. [18] segmented various types of lung nodules with fuzzy connectedness analysis, and Goncalves' team [19] proposed a comprehensive Hessian-based strategy that utilizes the concepts of central medialness adaptive principle, shape index and curvedness.

With the recent advancements in Machine Learning, techniques from this field have also been applied to solve the lung nodule segmentation task. While there do exist attempts where simpler models such as support vector machine (SVM) [20] are used to solve the problem, to the best of our knowledge, Convolutional Neural Network (CNN) is by far the most frequently adopted class of models in nodule segmentation as well as in most other image-related fields. In the LUNA16 grand challenge [21] and the Data Science Bowl 2017 [22], two major events focusing on lung nodule analysis and lung cancer detection, numerous participants demonstrated dozens of different CNN models and multiple techniques to improve model performances. Follow-up studies are still being conducted even years after the competitions ended. For example, Dou's team [23] proposed a three-dimensional (3D) CNN network that extracts and combines textural features at three different magnification levels; Zhu's team [4] utilized two consecutive 3D dual path networks to perform nodule segmentation and cancer classification respectively, and in the most recent work we found, a conditional generative adversarial network (cGAN) was used to artificially generate realistic nodule counterfeits to balance the various nodule types that were unevenly represented in the original data [17].

## III. DATASET

Our 3D thoracic CT scan images came from the LUNA16 dataset [21], which is a selected subset of the Lung Image Database Consortium image collection (LIDC-IDRI) [24] such

that scans with slice thicknesses (along the vertical, head-to-foot direction) greater than  $2.5\text{mm}$  are excluded. A total of 888 unique full-lung CT scans are included. The dimension of those slices in the transverse planes (normal to the head-to-foot axis) are  $512 \times 512$  pixels, but the total number of slices along the vertical direction, as well as the geometric scaling along the three Cartesian axes may vary among scans.

When preparing the nodule ground truths, instead of using the location-plus-approximate-radii ( $x, y, z, r$ ) annotation given by LUNA16, we used the voxel-level binary nodule masks available on the LIDC-IDRI website provided by 4 experienced radiologists according to the standards specified in Amato et al [25].

#### IV. METHOD

We approached the lung nodule segmentation task in a two-step process, allocating one module for lung segmentation and another for nodule segmentation. Due to the absence of ground truth in the lung segmentation step, lung segmentation was done using traditional image processing rather than a supervised network. As for the nodule segmentation module, we are aware that 3D CNN models outperform their 2D counterparts with other conditions being equal because they better incorporate equally-valuable information in all three Cartesian dimensions. Unfortunately, under the time constraint of the course project we decided to exclusively experiment with 2D CNN architectures for nodule segmentation. In the following subsections, we will go through the sequence of operations our framework performs, while a complete view of the framework will be provided at the end of this section.

##### A. Lung Segmentation

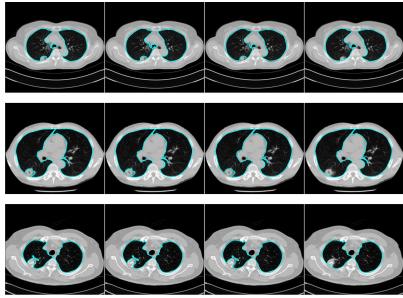


Fig. 1. An incomplete snapshot of Figure 1 in [1]. The images shown are the lung segmentation results of three cases using four different methods, with the last column being the "Active Contour with Bayesian Correction" method they proposed in the paper.

Using the idea and some source code provided by Chung et al. [1], we developed a traditional image processing module. The main reason we used the "Active Contour with Bayesian Correction" approach is that Chung et al. compared their method with some other existing ones and demonstrated its great performance on segmenting the juxtapleural nodules that reside on the lung boundaries as part of the lungs rather than as tissues outside the lungs. Here we included Figure 1, which

is an incomplete snapshot of a figure in their paper [1] to supplement their claims.

Our lung segmentation module runs the following procedure. Steps 2 – 4 are called preprocessing, while steps 5 – 7 are called lung segmentation.

- 1) Take in the 3D CT scans in MetaImage format (\*.mhd and \*.raw) [26] with intensity values in Hounsfield units.
- 2) Re-sample the 3D data to a consistent isotropic spacing of  $1 \times 1 \times 1$  mm among adjacent voxels.
- 3) Apply a global threshold at an experimentally-determined level at  $-800$  to merge the regions outside the human body.
- 4) Perform an intensity transform to match the standard for lung CT images found on Radiopaedia [27] using a window center at  $-600$  and a window width of 1500.
- 5) Segment the volumes that belong to the lungs using "Active Contour followed by Bayesian Corrections".
- 6) Remove everything except for the two largest 3D connected components in the scan. Also remove the second largest component if their volumes differ by more than 5 times or if the second largest component is smaller than 0.2 Liters. The ratio and volume thresholds were determined based on our basic knowledge about human lungs.
- 7) Pad all regions other than the lungs using a value of 45 which corresponds to the Hounsfield unit intensity of healthy muscle tissues, and
- 8) Store the 3D preprocessed (PP) and lung-segmented (LS) scans as DICOM (\*.dcm) [28] files for easy access in the following steps.

Figures 2 demonstrate our lung segmentation procedure summarized above.

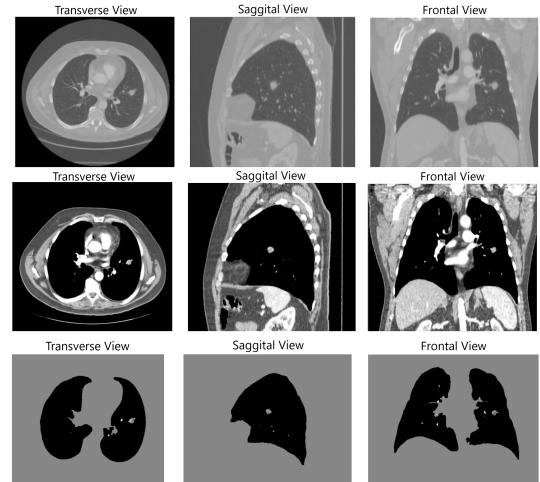


Fig. 2. Effect of Preprocessing and Lung Segmentation.

##### B. Lung Nodule Ground Truth (GT) Generation

The MATLAB toolbox shared by TesterTi [29] helped us generate the expert-annotated 2D pixel-level nodule maps

where 1's stand for nodule-related and 0's stand for non-nodule in every slice where one or more radiologists believed at least one nodule exists. Then we allocated the annotated slices back to where they belong in the 3D scans and created the ground truth for lung nodule segmentation in the form of 3D binary nodule masks. The same isotropic re-sampling was applied to match the preprocessed and lung-segmented scans. This process resulted in 4 ground truth scans for every lung-segmented scan, each contributed by one radiologist.

For consistency, we further integrated these 4 sets of ground truth scans into a single set using a rule such that for every voxel in a scan, its final binary value was set to 1 if it was annotated by at least 2 out of the 4 radiologists and 0 otherwise. The ground truths mentioned after this point were all referring to this set of scans we generated.

### C. Lung Nodule Segmentation

Under the help of the skeleton code provided by LeeJun-Hyun [30], we implemented several variants of U-Net for nodule segmentation. We split the lung-segmented (LS) scans into train, validation and test sets at a 8:1:1 ratio, and prepared the inputs to the nodule segmentation deep learning network in appropriate forms. The rationale behind randomization on the scan level is that we believed it was unreasonable to train and test the network using the images from the same patient. We also prepared a separate set of train, validation and test data using the corresponding preprocessed (PP) version rather than the lung-segmented versions of the same scans so that we could train the same network architectures on both sets to illustrate the effect of performing the lung segmentation procedure on the nodule segmentation result.

*1) Initial Slice-Based Model:* We initially attempted to perform the segmentation on one entire transverse slice at a time, but the preliminary segmentation results were poor, with most predictions for nodule-containing slices being all 0's. We attributed this phenomenon to the lack of data augmentation and the bad choice of loss function. Since the occurrences of nodules is rare by nature, with no more than a handful of them per 3D scan and they each cover a tiny fraction of any transverse 2D slice, thus the ground truth scans were heavily biased toward pixel values of 0's that correspond to non-nodules. Under such circumstances, predicting all pixel values to be 0's will yield a very high score as judged by the binary cross entropy loss function we used by default, and therefore the model was not motivated to correctly segmenting the nodules at all.

*2) Patch-Based Model and Data Augmentation:* Within any scan, the total count of nodule-related voxels is minimal. The percentage of nodule-related pixels is still overwhelmingly small even when we discarded the all-0's slices and only used transverses slices with nodules. To promote the representation of nodule-related pixels within each sample, instead of using the entire slices, we decided to extract regions of size  $32 \times 32$  around each nodule, where the  $(x, y, z, r)$  information from LUNA16 annotation defined the nodule locations. For each nodule, five patches were generated, and each patch was

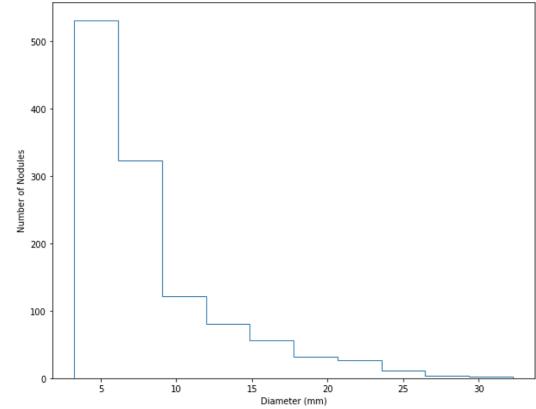


Fig. 3. Distribution of Lung Nodule Sizes in the LUNA16 Dataset.

randomly shifted in the x- and y- directions for any random integer value between -16 and 16. Since the orientation and voxel spacing are fixed among these scans and such information has medical implications, we didn't consider rotation, re-scaling, or flipping as appropriate data augmentation methods. Given the nodule size distribution as shown in Figure 3, by choosing such patch dimension and amount of shifting as mentioned, it is now possible for the fraction of nodule-related pixels in a patch to fall anywhere between 0 percent and 100 percent.

To prevent heavy bias toward nodule-related cases in the augmented dataset, within each LS or PP scan, we also randomly sampled 1 patch for every 4 patch generated using the method discussed above. These random samples may cover either nodule-free areas of the lungs, or areas outside the lungs. At a very low probability they may contain nodules as well. With such patch-based dataset preparation and data augmentation techniques, we had a reason to believe we reduced the bias in the dataset, and this was also supported by the results. The final LS and PP patch-based datasets prepared each contains 23,647 patches in the train set, 3,737 in the validation set and 3,002 in the test set. Note that this ratio slightly deviated from 8:1:1 because different scans do not necessarily contain the same number of nodules.

Since the current patch-based model only works with image patches of  $32 \times 32$  pixels, we would need a way to apply the patch-based model on  $32 \times 32$  regions over all transverse slices in a scan to be analyzed. The traditional way is to use a sliding window of dimension  $32 \times 32$  and step of 1 that iterate over the entire scan, and since each voxel was assigned a prediction score  $32 \times 32 = 1024$  times, the final segmentation result would be given by averaging these 1024 values and rounding that into a binary score. Instead of using this approach, we are experimenting with a more computationally favorable approach called nodule proposal network that can be found in the Discussion section. Before that is fully implemented, we would only test the patch-based nodule segmentation on  $32 \times 32$  regions rather than over entire scans.

*3) Choice of Loss Function:* The choice of loss function significantly impacts the segmentation performances. Even af-

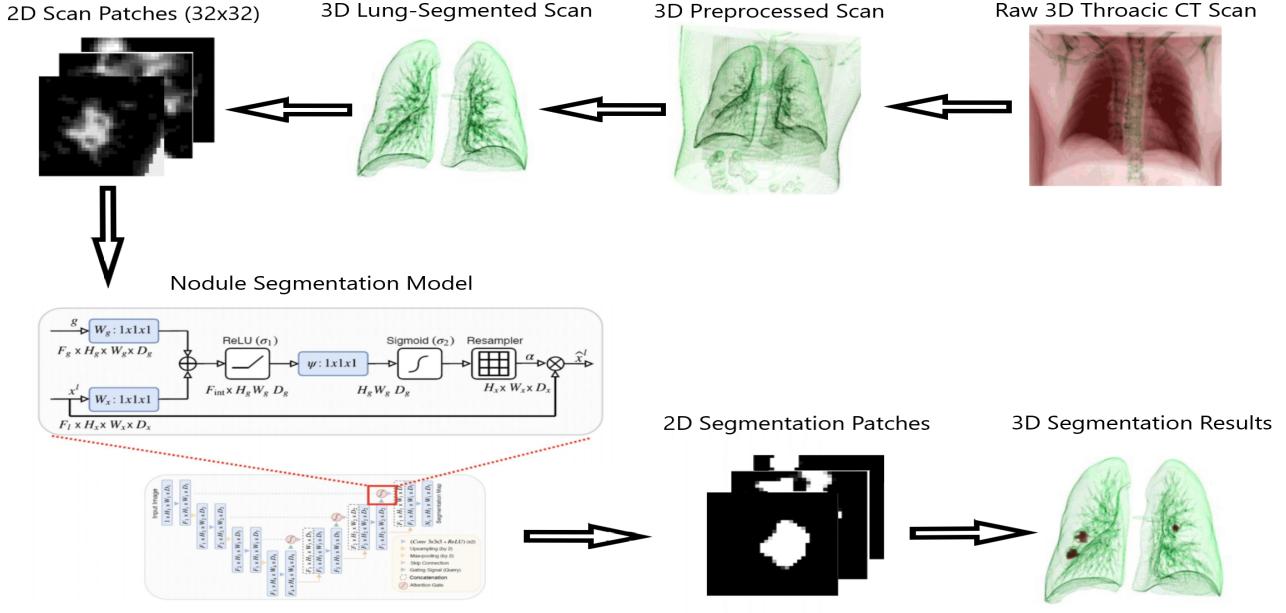


Fig. 4. Our Full Framework that Performs Lung Nodule Segmentation.

ter the patch-based dataset preparation and data augmentation, the total quantity nodule-related pixels were still fewer than non-nodule pixels, and it was still favorable for the network to predict all-0's under the binary cross entropy criterion. To promote the model's sensitivity to nodules, we used the dice coefficient loss with a smooth term:

$$\text{Dice\_Coef\_Loss} = -\frac{\text{sum}(\text{Prediction} \cap \text{GT} + \text{smooth})}{\text{sum}(\text{Prediction} + \text{GT} + \text{smooth})}$$

Since the Prediction and GT are both matrices (or vectors if flattening is applied) of the same dimension, the dice coefficient loss is effectively computing the intersection over union between the predicted image and the ground truth. The smooth term was set to 1 in our cases.

While the dice coefficient loss did encourage the model to detect the nodules and generate non-all-0's predictions when the ground truth contains nodules, it is still not ideal. Practically speaking, the quantitative scores given by the dice coefficient did not fully reflect our subjective judgement on the segmentation results, and from the theoretical standpoint, this loss function does not penalize the model from making arbitrary predictions on a slice where the ground truth is almost all-0's. We would like to modify this loss function so that it forces the model to behave reasonably both for nodule-present patches and nodule-absent patches.

Our proposed loss function, called the RR (Rosalie and

Raphael) dice coefficient loss, is given by:

$$\begin{aligned} \text{RR\_Dice\_Coef\_Loss} = & \\ & -\frac{\text{sum}(\text{Prediction} \cap \text{GT})}{\text{sum}(\text{Prediction} + \text{GT})} \\ & -\lambda \frac{\text{sum}(\overline{\text{Prediction}} \cap \overline{\text{GT}})}{\text{sum}(\overline{\text{Prediction}} + \overline{\text{GT}})} \end{aligned}$$

By fine-tuning the inverse ratio  $\lambda$ , we could balance the strength of its regulation on 1-valued predictions and 0-valued predictions.

*4) Network Architectures Assessed:* We assessed four variants within the U-Net family, namely U-Net [5], Attention U-Net (AttU-Net) [2], Recurrent Residual U-Net (R2U-Net) [6], and Attention Recurrent Residual U-Net (AttR2U-Net), with the last being the combination of AttU-Net and R2U-Net.

#### D. Full Framework

Our full framework is shown in Figure 4 on the top of the page. In the figure, we used the LS scan patches and the Attention U-Net to respectively represent the input and architecture of the nodule segmentation model because we later concluded that this combination had the best overall performance. In reality, the 2D scan patches generated from the LS scans as well as from the PP scans were combined with the 4 different network architectures and their performances were compared in the Results section.

To compare the performances of the 4 network architectures in a meaningful manner, we kept all of them at their default configurations as delineated in [30], with the exception that we changed the input tensor from 3 channels to 1 channel as our image patches were all in gray-scale, resulting in a input shape of  $(1, 1, 32, 32)$ .

TABLE I  
PERFORMANCES OF THE EIGHT ARCHITECTURE-SCAN SOURCE PAIRS

Architecture	Scan type	Accuracy	Sensitivity	Specificity	F1 Score	Jaccard Similarity	Dice Coefficient	RR_DC	inverse_ratio
R2AttU_Net	PP	0.71	0.43	0.700	0.380	0.285	0.214	0.295	0.506
R2AttU_Net	LS	0.80	0.54	0.640	0.422	0.398	0.339	0.408	0.540
AttU_Net	PP	0.76	0.42	0.768	0.410	0.364	0.296	0.383	0.568
AttU_Net	LS	0.77	0.34	0.782	0.339	0.305	0.249	0.331	0.543
U_Net	PP	0.76	0.32	0.768	0.362	0.292	0.229	0.315	0.530
U_Net	LS	0.75	0.36	0.764	0.337	0.294	0.236	0.316	0.532
R2U_Net	PP	0.74	0.41	0.736	0.422	0.306	0.227	0.318	0.528
R2U_Net	LS	0.73	0.67	0.549	0.345	0.358	0.300	0.367	0.485

## V. RESULTS

The 4 candidate network architectures paired with the 2 scan sources (LS versus PP) yielded 8 unique combinations. We needed to select the best network architecture and scan source in a justifiable manner.

1) *Performances of Different Network Architectures:* We exhausted all these eight possible combinations and constructed Table I. The "inverse\_ratio" in the last column of the table is the parameter  $\lambda$  in the RR dice coefficient loss function, which is not applicable in these cases as all of them were trained with the regular dice coefficient loss with smooth term.

These statistics alone did not provide enough information. It would have appeared that the Recurrent Residual U-Net paired with the Lung-Segmented scans yielded the best performance, but we quickly realized that was not the case based on further analysis.

A quantitative measurement other than the evaluation metrics to consider is the train and validation loss history shown in Figure 5. The number on the horizontal axis indicated the batch indices, each of which corresponding to 2,000 training patches as we set the batch size to 1 and ran the validation every 2,000 batches. These loss histories demonstrated signs of overfitting even before the training set had been exhausted. We would argue that such behaviors reflected a common constraint faced by all eight combinations, which is the fact that the ground truth annotations were not made with perfection by the radiologists. With the ground truths being significantly non-ideal, the simpler architectures yielded smoother validation loss and consequently, better segmentation results which will be further illustrated.

A subjective yet more intuitive way to analyze the segmentation performances is to look at some sample results. The example patch in Figure 6 pinpoints a significant problem with R2AttU-Net and R2U-Net, which was already mentioned as we discussed the choice of loss function. As the dice coefficient loss function has weak regulation on the segmentation prediction when the ground truth patch is all-0's, and the two more complicated models are more likely to overfit when the ground truth is not perfectly prepared, they happened to be not as well-behaved as their simpler variants in the same U-Net family.

By comparing the objective metrics and subjective visual inspection, we concluded that the two most meaningful metrics were the regular dice coefficient and specificity. The former

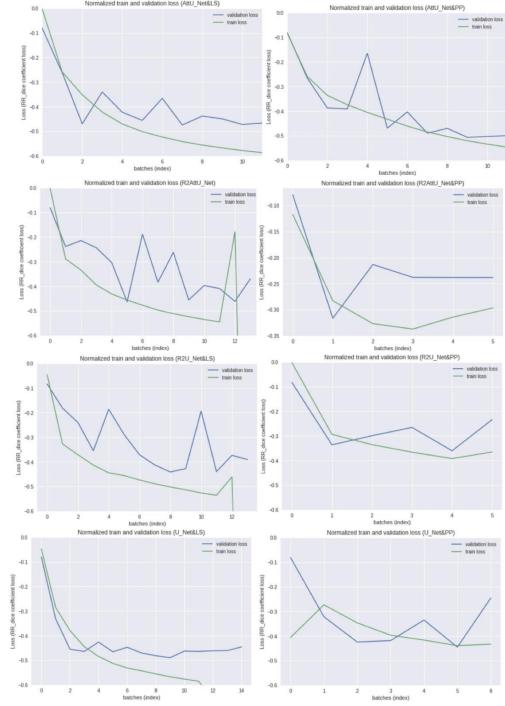


Fig. 5. Train and Validation Losses of the Eight Architecture-Scan Source Pairs

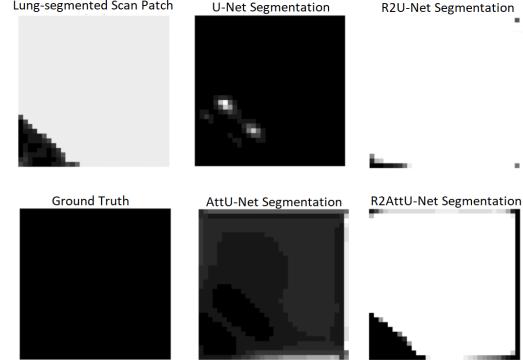


Fig. 6. Segmentation Results Produced by the Four Models Using Different Network Architectures and the Lung-Segmented (LS) Scans

effectively computes the intersection of union between the predicted and ground truth nodule areas, but suffers from its lack of regulation power when the ground truth is all-0's, while the latter indicates the false positive rate which largely complements the disadvantage of the former.

Based on the quantitative and qualitative analysis results, along with our intuition of the specializations of these network architectures, we eventually chose Attention U-Net (AttU-Net) as the most suitable architecture. Starting at this point, we focused on AttU-Nets exclusively and all further studies were conducted only on this architecture.

2) *Effects of the Loss Function:* The segmentation performances of the AttU-Net models trained with the regular dice

TABLE II  
PERFORMANCES OF ATTU-NET WITH DIFFERENT LOSS FUNCTIONS AND SCAN SOURCES.

Architecture	Scan type	Accuracy	Sensitivity	Specificity	Precision	F1 Score	Jaccard Similarity	Dice Coefficient	RR_DC	inverse_ratio
AttU_Net	PP	0.76	0.42	0.768	0.410	0.364	0.296	0.383	0.568	NaN
AttU_Net	LS	0.77	0.34	0.782	0.339	0.305	0.249	0.331	0.543	NaN
AttU_Net	PP	0.81	0.43	0.829	0.485	0.442	0.378	0.436	0.441	0.10
AttU_Net	LS	0.85	0.46	0.830	0.487	0.452	0.381	0.452	0.476	0.10
AttU_Net	PP	0.77	0.39	0.780	0.435	0.352	0.275	0.371	0.437	0.25
AttU_Net	LS	0.76	0.35	0.768	0.341	0.296	0.235	0.316	0.391	0.25
AttU_Net	PP	0.76	0.47	0.766	0.407	0.369	0.294	0.385	0.503	0.50
AttU_Net	LS	0.76	0.27	0.786	0.348	0.267	0.209	0.293	0.437	0.50
AttU_Net	PP	0.77	0.38	0.782	0.484	0.359	0.275	0.377	0.538	0.75
AttU_Net	LS	0.75	0.37	0.762	0.350	0.302	0.235	0.321	0.501	0.75

coefficient loss function as well as with our proposed RR dice coefficient loss function at various levels of the parameter  $\lambda$  are summarized in Table II. The two most meaningful metrics were also plotted in Figure 7. Based on similar quantitative and qualitative judgements, we would conclude that  $\lambda = 0.1$  yielded the best performance. One important thing to emphasize is that we should not blindly believe in the quantitative metrics, especially when the ground truths are far from ideal and the evaluation metrics are not fully informative.

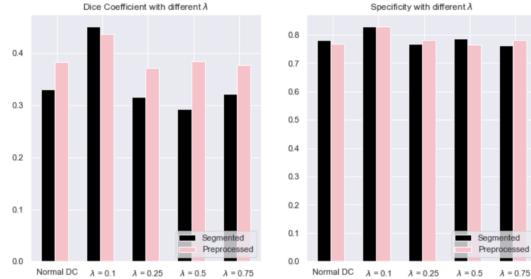


Fig. 7. Segmentation Results Produced by AttU-Net with Different Loss Functions and Scan Sources.

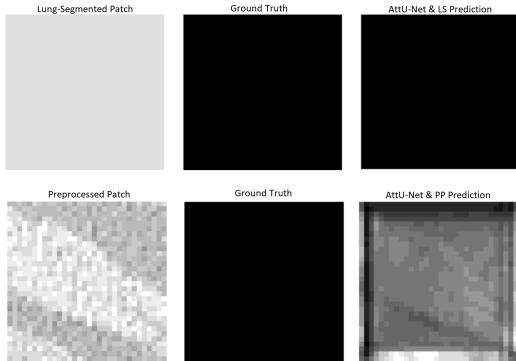


Fig. 8. Segmentation Results Produced by AttU-Net with Different Scan Sources on the Same Nodule-Absent Patch.

3) *Necessity of Lung Segmentation:* The lung-segmented scans did help produce better quantitative and qualitative segmentation results for various reasons. The most obvious cases were with the patches sampled from regions outside the lungs, since these volumes were padded with a uniform value in LS scans, it is much easier for the segmentation network to

learn that those regions were nodule-absent, as illustrated in Figure 8. Furthermore, in cases where the lung segmentation step nicely segmented juxtapleural nodules, the model with LS scans justifiably performed better than the model with PP scans, and even in cases where the lung segmentation step failed to include the juxtapleural nodules as part of the lungs, the nodule segmentation result given by the two models were still comparable. In summary, the model with LS scans was favored.

4) *Selected Segmentation Results:* Here we include some segmentation results of our final selected model (AttU-Net + LS scan patches) to subjectively showcase its performance. The network indeed perform sensible segmentation and in some cases the segmentation is more reasonable than the non-ideal ground truth.

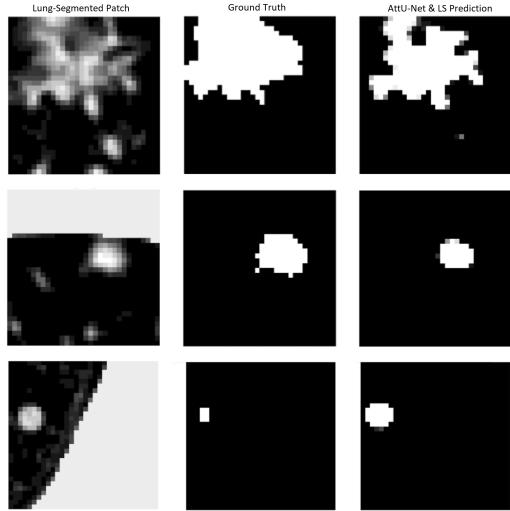


Fig. 9. A Selection of Segmentation Results Produced by the Model Using AttU-Net with Lung-Segmented Patches.

5) *Visualizations of the Convolutional Layers:* In our attempt to explore what the deep learning network was learning and to make sense of its decisions, we adopted an approach where we demonstrated the behaviors of the convolutional filters by artificially generating the input images that would maximize their responses. These images, sometimes called activation maps, were generated through a process called gradient ascend, where the pixel-level intensity values of the images were adjusted to invoke the highest level of response on the corresponding convolutional filter. By observing these images we could get a sense of what patterns and textures each convolutional filter is most sensible to. That was not the most straightforward representation of the knowledge acquired by the model, but at least it could provide some insight.

Figure 10 shows the images that generated the strongest responses to the first 16 convolutional filters in the first among the five convolutional blocks in the Attention U-Net architecture, while Figure 11 shows their counterparts for the last convolutional block. The convolutional filters are not sorted in any sense for any convolutional block, and therefore

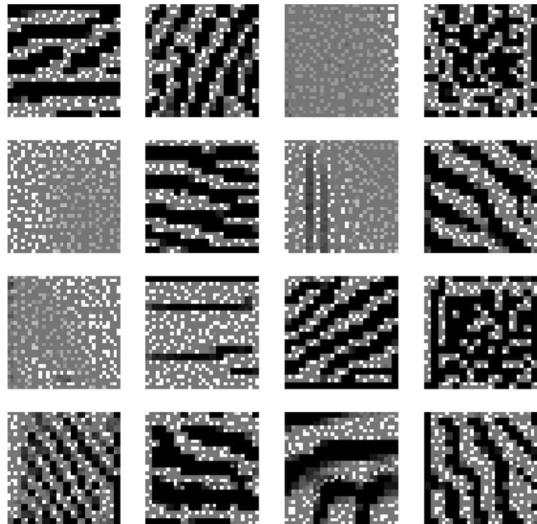


Fig. 10. Images that Generate the Strongest Responses in the First 16 Convolutional Filters in Convolutional Block 1.

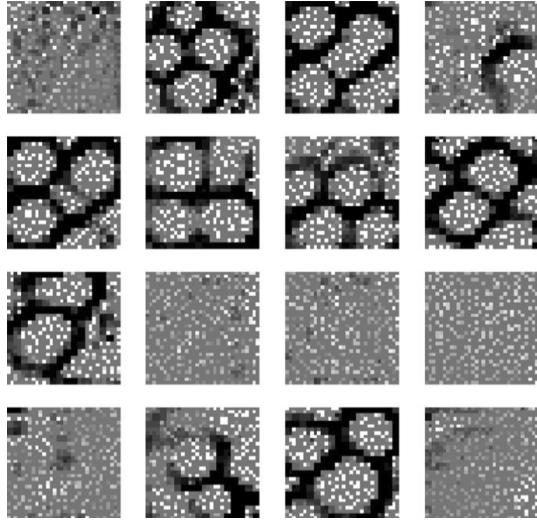


Fig. 11. Images that Generate the Strongest Responses in the First 16 Convolutional Filters in Convolutional Block 5.

the first 16 such filters are mathematically equivalent to 16 random filters.

As we could see by comparing these two figures, the activation maps of the filters in a shallower convolutional block possessed simpler patterns and textures, such as alternating intensities along the horizontal, vertical and diagonal directions. The activation maps deeper down the network, however, had more sophisticated patterns. For our specific application, since the deep learning network was trained to identify the lung nodules, which are mostly circular in shape, it makes perfect sense that the deeper convolutional filters were sensitive to circular patterns that resembled the size and shape of the lung nodules.

## VI. CONCLUSION

We successfully implemented the lung nodule segmentation framework that could produce visualizations of lung and nodule segmentation. Figure 12 is an example of our visualization result where the volumes corresponding to the lungs and lung nodules were stained in distinct colors. We experimentally determined that Attention U-Net along with the Lung-Segmented Scan Patches yielded the best performance among the combinations assessed. We also found the most meaningful evaluation metrics are the regular dice coefficient and specificity for reasons discussed in the Results section. Finally we confirmed that our proposed RR dice coefficient loss function with  $\lambda = 0.1$  did improve the segmentation performance.

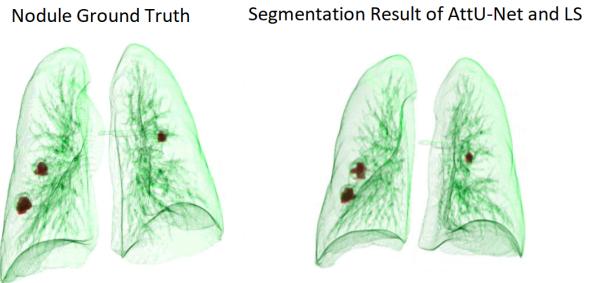


Fig. 12. Visualization of a Sample Segmentation Results Produced by the Model Using AttU-Net with Lung-Segmented Patches.

## VII. DISCUSSION

*1) Limitations of the Active Contour with Bayesian Correction Model:* Unlike what Chung et al. advertised in their paper, the lung segmentation performed by the "Active Contour with Bayesian Correction" approach was quite problematic, and in fact it only properly handled less than a quarter of juxtapleural nodules, besides having the tendency of incorrectly segmenting tissues away from the lungs. For a better lung nodule segmentation framework, we would like to improve the lung segmentation results using other methods accompanied with manual corrections, and eventually delegate such jobs to deep learning models as well.

*2) Imperfections in Nodule Ground Truth Annotations:* The ground truth annotations given by the 4 radiologists turned out to be far from perfect as well. Similar to what we need to do to handle the imperfections in lung segmentation, we would have to consult experts and potentially perform manual fixes to polish the current annotations.

## VIII. NEXT STEPS

Besides dealing with the issues mentioned in the Discussion section, we would also perform the two following modifications on our framework.

*1) Nodule Proposal Network:* In order to perform whole-scan segmentation with the patch-based model, we are trying to employ a deep learning network to generate potentially nodule-containing  $32 \times 32$  patches for the model to analyze.

The Faster R-CNN model from [4] was designed for almost exactly this purpose, with the only distinction that the patch proposals generated are 3D blocks of dimension  $32 \times 32 \times 32$ . To accommodate this dimension conflict, we could either perform the segmentation on 2D patches and iterate the process the 32 times, or, as a direction we would pursue, we could upgrade our nodule segmentation model from 2D to 3D.

*2) Upgrading the Nodule Segmentation Model from 2D to 3D:* The only advantage of a 2D convolutional neural network to its 3D counterpart is the relatively lower computational cost and hence the time it takes to implement the model. Since the latter makes use of the previously unused information from the third dimension, we can reasonably expect with a high confidence that the reliability of the nodule segmentation model shall increase as we upgrade it to its 3D version. The modification is almost as simple as modifying the shape of the input tensor to the Attention U-Net model from  $(1, 1, 32, 32)$  to  $(1, 32, 32, 32)$  and adjusting the output shape accordingly.

## IX. ACKNOWLEDGMENT

The authors would like to thank the course instructors, Professor Andrew Laine and Professor Paul Sajda for the instruction and guidance, the great facilitation and support provided by Professor Jia Guo outside the class, and academic advisor Professor Christine P. Hendon for the help related to the image processing aspects.

## REFERENCES

- [1] Chung, Heewon, Hoon Ko, Se Jeong Jeon, Kwon-Ha Yoon, and Jinseok Lee. 2018. "Automatic Lung Segmentation With Juxta-Pleural Nodule Identification Using Active Contour Model and Bayesian Approach." *IEEE Journal of Translational Engineering in Health and Medicine* 6 (May): 1800513.
- [2] O. Oktay et al. (2018). Attention U-Net: Learning Where to Look for the Pancreas. [online] arXiv.org. Available at: <https://arxiv.org/abs/1804.03999v3> [Accessed 9 May 2019].
- [3] Key Statistics for Lung Cancer. [Online]. Available: <https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html>. [Accessed: 07-May-2019].
- [4] W. Zhu, C. Liu, W. Fan, and X. Xie. 2018. DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). <https://doi.org/10.1109/wacv.2018.00079>.
- [5] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Lecture Notes in Computer Science. pp. 234241, 2015.
- [6] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari. 2019. "Recurrent Residual U-Net for Medical Image Segmentation." *Journal of Medical Imaging* (Bellingham, Wash.) 6 (1): 014006.
- [7] M. Kaneko et al., "Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography," *Radiology*, vol. 201, no. 3, pp. 798802, Dec. 1996.
- [8] P. B. Bach et al., "Benefits and harms of CT screening for lung cancer: a systematic review," *JAMA*, vol. 307, no. 22, pp. 24182429, Jun. 2012.
- [9] Z. Saghir et al., "CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT," *Thorax*, vol. 67, no. 4, pp. 296301, Apr. 2012.
- [10] T. B. Lanni et al., "Early Results From the Implementation of a Lung Cancer Screening Program," *American Journal of Clinical Oncology*. p. 1, 2015.
- [11] T. N. L. S. T. R. Team and The National Lung Screening Trial Research Team, "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395409, 2011.
- [12] D. E. Wood et al., "Lung Cancer Screening, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology," *J. Natl. Compr. Canc. Netw.*, vol. 16, no. 4, pp. 412441, Apr. 2018.
- [13] J. H. Austin et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. In: *Radiology* 200.2 (1996), pp. 327331.
- [14] V. S. Nair, V. Sundaram, M. Desai, and M. K. Gould, "Accuracy of Models to Identify Lung Nodule Cancer Risk in the National Lung Screening Trial," *Am. J. Respir. Crit. Care Med.*, vol. 197, no. 9, pp. 12201223, May 2018.
- [15] I. Vlahos, K. Stefanidis, S. Sheard, A. Nair, C. Sayer, and J. Moser, "Lung cancer screening: nodule identification and characterization," *Transl Lung Cancer Res*, vol. 7, no. 3, pp. 288303, Jun. 2018.
- [16] J. Kalpathy-Cramer et al., "A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study," *J. Digit. Imaging*, vol. 29, no. 4, pp. 476487, Aug. 2016.
- [17] Y. Qin, H. Zheng, X. Huang, J. Yang, and Y.-M. Zhu, "Pulmonary nodule segmentation with CT sample synthesis using adversarial networks," *Med. Phys.*, vol. 46, no. 3, pp. 12181229, Mar. 2019.
- [18] P. Badura and E. Pietka, "Soft computing approach to 3D lung nodule segmentation in CT," *Comput. Biol. Med.*, vol. 53, pp. 230243, Oct. 2014.
- [19] L. Goncalves, J. Novo, and A. Campilho, "Hessian based approaches for 3D lung nodule segmentation," *Expert Systems with Applications*, vol. 61, pp. 115, 2016.
- [20] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: a complete intelligent system," *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287300, May 2013.
- [21] "LUNA16-Data." [Online]. Available: <https://luna16.grand-challenge.org/Data/>. [Accessed: 23-Mar-2019].
- [22] Data Science Bowl 2017. [Online]. Available: <https://kaggle.com/c/data-science-bowl-2017>. [Accessed: 08-May-2019].
- [23] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection, *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 15581567, Jul. 2017.
- [24] "LIDC-IDRI - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki." [Online]. Available: <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>. [Accessed: 23-Mar-2019].
- [25] S. G. Armato 3rd et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915931, Feb. 2011.
- [26] ITK/MetaIO/Documentation - KitwarePublic. [Online]. Available: <https://itk.org/Wiki/ITK/MetaIO/Documentation>. [Accessed: 08-May-2019].
- [27] A. Murphy, Windowing (CT) — Radiology Reference Article — Radiopaedia.org, Radiopaedia. [Online]. Available: <https://radiopaedia.org/articles/windowing-ct?lang=us>. [Accessed: 09-May-2019].
- [28] Softneta, DICOM Library - About DICOM format. [Online]. Available: <https://www.dicomlibrary.com/dicom/>. [Accessed: 08-May-2019].
- [29] TesterTi. [Online]. "TesterTi/LIDCToolbox." GitHub. <https://github.com/TesterTi/LIDCToolbox>. [Accessed April 22, 2019]
- [30] LeeJunHyun. [Online]. "LeeJunHyun/Image\_Segmentation." GitHub. [https://github.com/LeeJunHyun/Image\\_Segmentation](https://github.com/LeeJunHyun/Image_Segmentation). [Accessed April 22, 2019].