Lung Nodule Detection

Erica Chio EBC308@NYU.EDU
NYU Sackler Institute Masters Student

Nithu Mathew NJM363@NYU.EDU
NYU Sackler Institute Masters Student

Abstract

Lung cancer is one of the leading causes of cancer-related deaths due to its aggressive nature, but early detection of the cancer can greatly improve chances of survival. Typically, computed tomography (CT) scans are used to diagnose lung cancer through identification and classification of lung nodules. In this project, two deep learning models, Faster-RCNN and Mask R-CNN, were proposed for lung nodule detection and evaluated on CT scans from LIDC-IDRI datasets. The models were implemented using two pretrained ResNet-50 backbones, one trained on the Common Objects in Context (COCO) dataset and one trained on ImageNet. It was hypothesized that Mask R-CNN would do better than Faster-RCNN because of the additional implementation of segmentation masks. While both models performed equally in detecting all nodules with area greater than 100, Faster R-CNN did better at detecting nodules with area greater than 50. Both models performed better on the COCO dataset compared to ImageNet.

1. Introduction

Lung cancer is one of the major causes of cancer-related deaths due to its severity and delayed detections at advanced stages. Early detection of lung cancer is very important for the survival of an individual. Computed tomography (CT) scans are considered the most suitable in diagnosing lung diseases and thus, are used initially for the diagnosis of malignant nodules.[13;8] Lung nodules are spots that are more solid than normal lung tissue and will show up on X-rays or CT scans. These lung nodules are common, however, in some cases, these nodules can indicate early stages of lung cancer. Early stage cancerous nodules are quite similar to noncancerous nodules.[11] Thus, it is important for the sizes of these nodules to be tracked as time goes on, because growth of the nodule indicates potential lung cancer.[16]

There are different techniques to try and identify and classify these lung nodules, such as CT scans, which looks at the morphological aspect of the nodule, positron emission tomography (PET), which analyzes the metabolic aspect of the nodule, and needle prick biopsy analysis.[20] However, surgeries and biopsies are necessary to differentiate between benign and cancerous, adding a lot of stress to patients. Currently, results show that if only one radiologist reviews the CT scan, that only 68% of the time is the nodule diagnosed correctly. Accuracy goes up to 82% when two radiologists screen, demonstrating that nodule detection is time consuming and error prone.[4] Having a computer-aided diagnosis (CAD) can potentially help increase nodule detection rate. As such, having computer-aided diagnosis (CAD) to identify these nodules without requiring invasive surgery would help alleviate patient anxieties about surgery and reduce radiologists' workloads. CAD is already outperforming radiologists, but often has a high false positive rate.[9]

Deep learning algorithms, in particular convolutional networks, have become increasingly popular in the medical imaging field, including computer-aided diagnosis (CAD), radiomics, and medical image analysis.[17] CAD is commonly used for classification of objects.[17] For medical image analysis, detection and classification is performed by taking images as input and extracting features by partitioning the image into segments in order to concentrate on a specific area. The model typically used for image analysis is called convolutional neural network, or CNN.[12] A CNN has multiple layers of convolutions and activations, often interspersed with pooling layers, and is trained using backpropagation and gradient descent.[15] Another such class of machine learning is massive-training artificial neural network (MTANN), which when applied reduced false positives.[18]

In particular, a 3D Faster R-CNN model, used by Ding and Liao et al. was used to reduce false positives.[19] Faster-RCNN introduces the idea of Regional Proposal Methods (RPNs), which is the backbone of a Faster-RCNN model.[14] RPNs have both a classifier and a regressor. The classifier determines the probability of an object, while the regressor determines the coordinates. These proposals are generated by having a small sliding window slide over the last convolutional feature map. These proposals have an "anchor," which is the central point of the sliding window.[14]

Another proposed model for object detection is Mask R-CNN, which further extends Faster-RCNN by implementing a branch to detect segmentation masks on the Region of Interest (RoI) in addition to the existing classification and bounding box regression. The mask branch includes a fully convolutional network (FCN), which predicts segmentation masks for each RoI. The segmentation masks encode the input's object spatial layout, therefore can be addressed in a pixel-to-pixel manner, rather than being collapsed into short output vectors as is done for classification labels and bounding box coordinates.[7]

We propose using pre-trained Faster-RCNN and Mask R-CNN models to detect lung nodules.

2. Data

2.1 Data Source

The model is evaluated on a dataset from the Lung Image Database Consortium image collection (LIDC-IDRI) dataset, found in The Cancer Image Archive database. The LIDC-IDRI image collection is an open-access resource for development and evaluation of computer-aided diagnosis (CAD) methods to detect and diagnose lung cancer. Several academic centers and medical image companies collaborated to produce the complete image collection containing diagnostic and lung cancer computed tomography (CT) scans from 1010 sample cases and a total of 244,167 images.

In this dataset, the images have been marked up for lesions by 4 different radiologists who individually reviewed each image and annotated lesions belonging to one of three categories ("nodule > or =3 mm," "nodule < 3 mm," and "non-nodule > or =3 mm"). Each radiologist then reviewed their own annotations along with those of other radiologists to come to a final conclusion. Each sample contains the annotated DICOM images from the CT scans and an XML file which records the results of the annotation process by the radiologists. The data can be downloaded with the NBIA Data Retriever application.

2.2 Data Preparation

Processing DICOM scans differs from PNG and JPEG images since the medical imaging has a huge amount of data within the pixels and metadata. DICOM files have to be specifically loaded in with DICOM packages to ensure its pixel data and metadata is not lost. The pixels then had to be converted into Household Units, a measure of radiodensity. These units represent whether that area is flesh, bone, or air. DICOM images also differ from image to image in pixel spacing, meaning there had to be standardization and normalization among all the images. Without this standardization of pixel spacing, the model would have a hard time understanding the images. The mask for the Mask R-CNN, a binary image, also had to be created by segmenting the lungs, and some tissue around. The mask only kept the largest air pocket, labelling air as 0 and lung tissue as 1. This was done using connected component analysis.[3]

3. Materials and Methods

3.1 Transfer Learning Model

We used transfer learning to create our models. Both the Faster R-CNN model and Mask R-CNN model utilized Resnet-50 as its backbone. The architecture of ResNet-50 has 4 main stages. The network performs an initial convolution with max pooling before entering into the stages. Each stage consists of a number of residual blocks containing 3 convolutional layers each. As input progress from one stage to another, the channel width is doubled and the size of the input is reduced to half. The network ends with a global average pooling layer and a 1000-way fully connected layer with softmax.[6] In the implementation of the ResNet-50 backbone, layer 2, 3 and 4 had their weights frozen. A Feature Pyramid Layer was also added to the top of the model. They are required to help detect model of different scales, which is vital in our model as our bounding box areas ranged from 50 to 2000 pixels.[10]

3.2 Loss

Faster R-CNN and Mask R-CNN used a multi task loss function. Faster R-CNN used a combination of a classification and bounding box regression loss. The classification loss is a log loss function, which allows transition from multi class to binary classification. The bounding box regression loss is Smooth L1 loss.[5;14] Mask R-CNN also uses a multi task loss function, consisting of classification, bounding box regression, and mask loss. The mask loss is defined as an average binary cross entropy loss as it is trying to determine a mask for each class, so there is no multi class cross over.[7]

3.3 Evaluation Metrics

The performance of object detection classification models are generally evaluated using mean average precision, where the average precision of each class is averaged to get the mean of the entire model. However, since our model only has one class, we evaluate our model using average precision. This is a popular metric that is measured by computing the average precision value for recall value over 0 to 1. Precision measures the proportion of accurate predictions to total predictions, while recall measures the proportion of accurate

predictions to total number in the positive class. In the context of our project, there is only one class (nodule) so recall will be 1.

For our models, the accuracy of the predicted bounding boxes was determined by calculating IoU (intersection over union). This metric measures how much the predicted boundary overlaps with the ground truth. Predictions with IoU score above the threshold (0.5) were considered to be a positive prediction. The total number of positive predictions were averaged to calculate average precision for the model.

3.4 Hyperparameter Selection

- **Area of Bounding Box:** We played with the lower limit to the area of bounding box. The dataset included nodules $<$ 3mm, which in practice would be marked as a point. Thus, these nodules did not have any area within their bounding box. As such, we decided to remove these nodules, and only focus on nodules $\geq$ 3mm. These nodules area ranged from 1.0 pixel wide to 2000.0 pixels wide. From these nodules, we decided to try lower limits of 50 and 100. A larger area tended to do better, potentially, because range was smaller and the object was big enough to identify.

- **Pretrained Domain:** We sampled the same backbone trained on two different image data set, ImageNet and COCO. ImageNet has about 27 high level categories and 14 million images.[(2)] On the other hand, COCO has about 300,000 images and 91 categories.[(1)] We found that the backbone pretrained on COCO to have a better resulting test accuracy.

- **Learning Rate:** We tried two manual (StepLR) and automatic (ReduceLROnPlateau) learning rate schedulers from Pytorch. We found that StepLR led to a steady learning rate, and was easier to understand and implement than ReduceLROnPlateau.

- **BatchSize:** We sampled with different batch sizes: 1,4, and 8. A Stochastic Gradient batch size of 1 was too small, and created too much noise, resulting in lower performance. We also tested a minibatch size of 8, but found that the optimal minibatch size was 4, as it had the best performance.

4. Results

ResNet-50 as a backbone worked well regardless of which image dataset it was trained on, although COCO trained backbone did better in both Faster R-CNN and Mask R-CNN models. Testing the lower bounding box area for the models did not result in any huge difference, although for Faster R-CNN, it had a higher test accuracy with a lower bounding box area, while for Mask R-CNN, having a larger bounding box resulted in better test accuracy. Faster R-CNN and Mask R-CNN both performed well, with Faster R-CNN performing only a little bit better.

Validation accuracy curves can give us a better understanding of the model. Interestingly, the validation curves for ImageNet had more learning to do, but would still achieve the same accuracy at the end. All of the validation curves would plateau around 70% accuracy.

Model	Backbone	Transfer Learning	Area Included (≥)	Test Accuracy
Faster R-CNN	ResNet-50	COCO	50	78.0%
		COCO	100	77.2%
		ImageNet	100	73.0%
Mask R-CNN	ResNet-50	COCO	50	74.5%
		COCO	100	77.8%
		ImageNet	100	73.6%

Table 1: Test Accuracy for Each Model

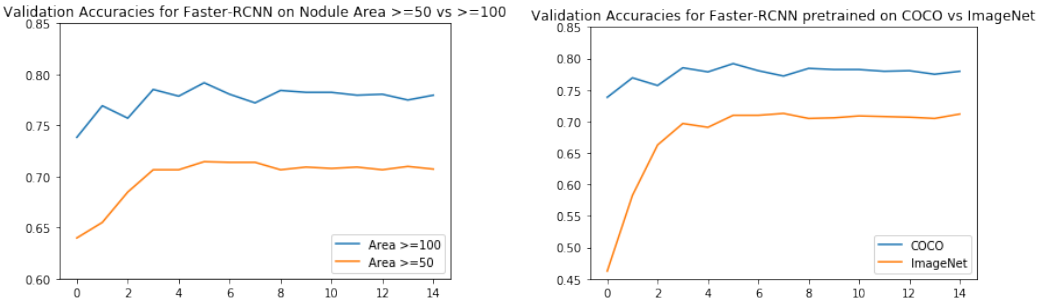


Figure 1: Faster R-CNN Validation Accuracy Curves

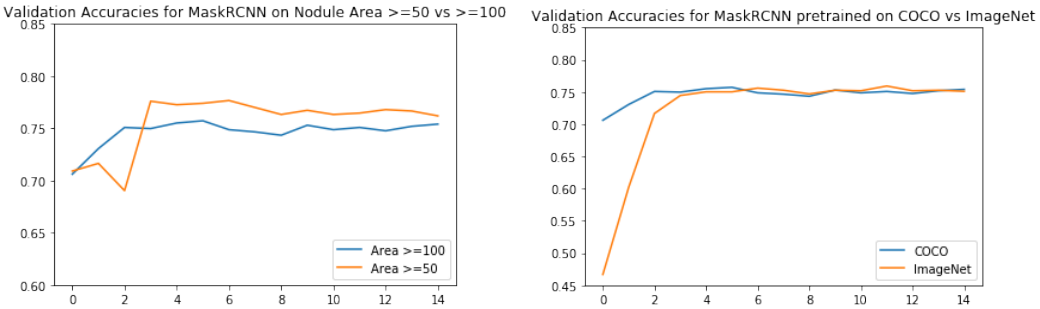
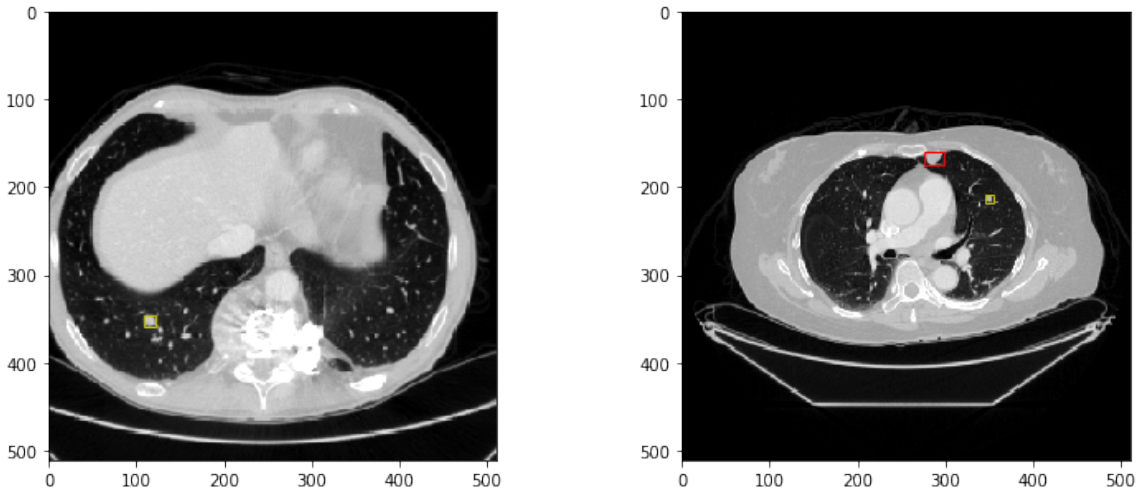


Figure 2: Mask R-CNN Validation Accuracy Curves

5. Discussion

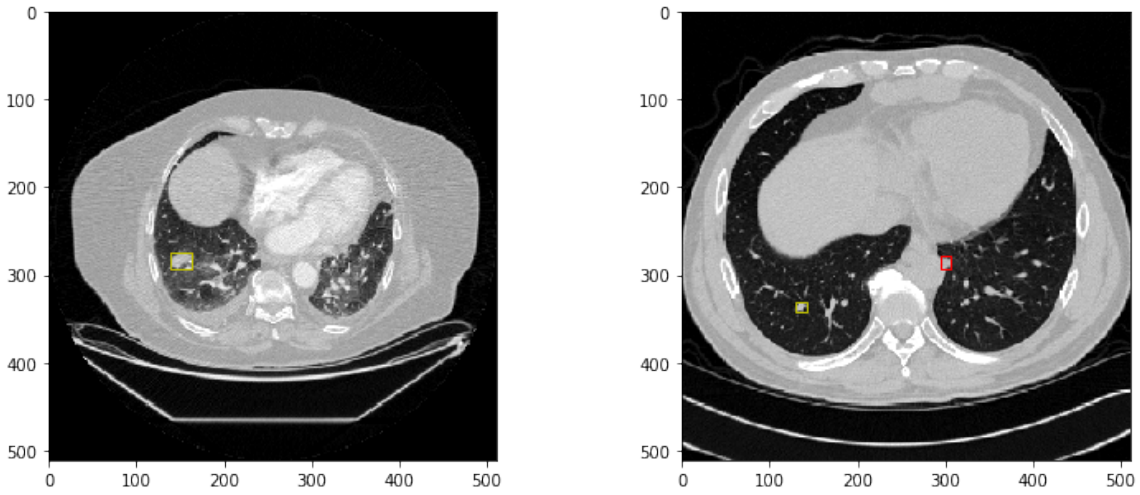
Although it was hypothesized that Mask R-CNN would perform better than Faster R-CNN, both models did equally well in detecting lung nodules. We have generated sample correct and incorrect predictions for both Faster R-CNN and Mask R-CNN models. For the correct predictions, the models performed well, with an IOU in the 0.90s. This is considered very good, as in practice an 0.5 IOU is considered correct. Interestingly, the incorrect predictions, had an IOU of 0.0. At first glance, it would seem like the model is completely incorrect about nodule identification. However, upon closer inspection, the model identifies something "nodule-like" as a nodule. The correct nodule is also no longer a "free standing" nodule, but rather a part of a bigger structure. This indicates the model

knows what a nodule is supposed to look like, but may have issues in identifying nodules that are a part of the bigger lung structure.



Correct Prediction with an IoU: 0.949 Incorrect Prediction with an IoU: 0.0

Figure 3: Faster R-CNN Predictions



Correct Prediction with an IoU: 0.987 Incorrect Prediction with an IoU: 0.0

Figure 4: Mask R-CNN Predictions

All code can be found on Github.

References

[1] URL `http://cocodataset.org/#home`.

[2] URL `http://www.image-net.org/about-stats`.

[3] URL `https://www.kaggle.com/akh64bit/full-preprocessing-tutorial`.

[4] Al Mohammad B., Brennan P.C., and Mello-Thoms C. A review of lung cancer screening and the role of computer-aided detection. *Clin. Radiol.*, 72(6):433–442, 2017.

[5] Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.

[6] K. He, X. Zhang, S. Ren, and Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[7] He K., Gkioxari G., Dollar P., and Girshick R. Mask r-cnn. *CoRR*, abs/1703.06870, 2017.

[8] Lee K.S., Mayo J.R., Mehta A.C., Powell C.A., Rubin G.D., Prokop C.M.S., and Travis W.D. Incidental pulmonary nodules detected on ct images: Fleischner. *Radiology*, 284 (1):228–243, 2017.

[9] Yu L., Dou Q., Chen H., Heng P.-A., and Qin J. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Trans. Biomed. Eng.*, 64 (7):1558–1567, 2016.

[10] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. *CoRR*, abs/1612.03144, 2016.

[11] Nair M. and Sandhu S.S.and Sharma A.K. Cancer molecular markers: A guide to cancer detection and management. *Seminars in Cancer Biol*, 52(1):39–55, 2018.

[12] Deepa N. and Chokkalingam. SP. Deep convolutional neural networks (cnn) for medical image analysis. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(3S):607–610, 2017.

[13] N. Nasrullah, Alam Sang, J., M. S., M. Mateen, B. Cai, and H. Hu. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors (Basel, Switzerland)*, 19(17):3722, 2019.

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015.

[15] L. Alexander Selvikvåg and L. Arvid. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

[16] R. S. Slatore, C. G.and Wiener and A. D. Laing. Patient education information series: What is a lung nodule? *American Journal of Respiratory and Critical Care Medicine*, 193(7):11–12, 2016.

[17] K. Suzuki. Overview of deep learning in medical imaging. *Radiol Phys Technol 10*, page 257–273, 2017.

[18] K. Suzuki, S. G. Armato III, F. Li, S. Sone, and K. Doi. Massive training artificial neural network (mtann) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. *Medical Physics*, 30(7):1602–1617, 2003.

[19] Zhu W., Liu C., Fan W., and Xie X. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification; proceedings of the ieee winter conference on applications of computer vision (wacv). *IEEE Trans. Biomed. Eng.*, 2018.

[20] Shi Z., Zhao J., Ji G. Han X.and Pei B., and Qiang Y. A new method of detecting pulmonary nodules with pet/ct based on an improved watershed algorithm. *PLoS ONE*, 2015.