Lung Cancer Detection with
Region Enhanced Multiple Instance Networks

Jay DeStories, Jason Fan, Alex Tong

March 2017

1 Introduction and Problem Statement

The problem of image segmentation and structure annotation within the field of biomedical imaging has become a well developed and very active field in the past years. In 2016 and 2015, the LUng Nodule Annotation (LUNA) challenge and SPIE Lungx challenge, asked researchers to develop models to identify pulmonary nodules in lung CT slices. With the 2016 LUNA Challenge, researchers gained access to annotated CT slices that contained segmentation ground truths for abnormal nodules but did not release data about the malignancy of the nodules. In the 2015 SPIE less than 80 CT annotated images of malignant nodules were released to the public.

Finally, with the 2017 Kaggle Data Science Bowl, a larger dataset of 1000+ lung CT images in DICOM format was finally released with cancer/no cancer labels. This has allowed researchers to answer a deeper question about lung CT scans; whether or not there are indicators of malignancy and cancer in a patient's CT scan.

There is, however, one caveat to the Kaggle dataset. Although the presence of malignancy is indicated by global, binary cancer/no-cancer label, the location of malignant nodules and structures are *not* annotated in the training data.

Inspired by recent work in biomedical image segmentation [1], region proposal networks and multiple instance learning for whole mammogram classification [2], We seek to present a novel pipeline for lung cancer detection that enhances multiple instance learning with region proposal.

2 Related Work

Over the past decade there has been a significant amount of work towards computer aided diagnosis of lung cancer [3]. Depending on what kind of data researchers have had access to, previous efforts to identify lung cancer can be categorized by two main approaches.

(1) Pulmonary Nodule Detection methods use image processing techniques to segment and annotate nodules [4, 5, 6]. These methods mimic radiologists by looking for abnormalities in the form of "solitary white nodule-like blob[s]" in a chest x-rays and CT scans. Lung nodules are potential cancer indicators, and as such are an important part in early lung cancer diagnosis. However, not all pulmonary nodules are malignant and many are benign, there will be false positives in diagnosis if we naively associate nodule presence with cancer.

(2) Direct inference and classification methods instead attempts to directly predict the probability of cancer using x-ray and CT images without nodule detection [7, 8].

One challenge that almost all researchers in biomedical image inference face is the problem with datasets being small and weakly labeled. Images labeled with cancer/no-cancer binaries are weakly labeled because imaged tissues only display malignancy locally; not *all* of the tissue in an image will have cancer. Multi-Instance Networks have been trained from binary labels to classify whole mammogram images [2].

2.1 Reference Papers

- Automatic Liver and Lesion Segmentation in CT using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields [1],

- Deep Multi-Instance Networks with Sparse Label Assignemnt for Whole Mammogram Classification [2],

- U-Net: Convolutional Network for Biomedical Segmentation [6].

3 Method/Algorithm/Pipeline

3.1 Architecture Overview

Our Lung Cancer detection method uses nodule annotation to enhance direct classification.
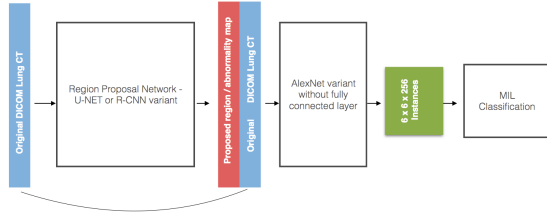
We plan to leverage state of the art region proposal and biomedical segmentation to indicate nodule presence and implicitly weight instances consumed by a downstream Multi-Instance Network (MIN) classifier.

In our pipeline, a region proposal network is trained on the LUNA dataset to propose nodule annotations and abnormal regions in CT slices. This region proposal is then used to annotate nodules in Kaggle dataset. Then a MIN classifier based on the work on mammogram classification by Lou et. al [9], will be trained on the annotated Kaggle dataset.

3.2 Training

We will train the upstream and downstream networks seperately. Once the upstream, region proposal/segmentation network is trained, we will then replace the last layers of the upstream network with convolutional layers to learn the features that best associate region proposals with malignancy.

3.3 Choosing the Region Proposal Network

We have several region proposal networks we want to investigate. We will tune one segmentation/region proposal network for our final pipeline. the region proposal networks that we suspect will yield promising results can be classified int two categories.

First, there exists segmentation networks based of U-Net, a seminal CNN for biomedical segmentation developed by Brox et. al [6]. The most easily accessible U-Net based pipeline for supervised biomedical image segmentation is a pre-trained network, available on Caffe Model Zoo, for CT Liver Lesion segmentation [1].

Second, there exists other highly successful, and fast, region proposal and segmentation networks developed for non-biomedical domains. We will test the fully convolutional neural network developed by Darrell et. al at Berkeley [10] (Jay and Jason will be investigating this network as a paper presentation). We will also test and try to adapt Faster R-CNN developed by He et. al [11, 12] for nodule detection.

Recent U-Net based segmentation leverage depth information whilst non-biomedical region proposal and segmentation networks do not use depth information. If time permits or if we deem necessary, we will attempt to adapt successful 2D methods to work in 3D.

3.4 The Multi-Instance Network

We will adapt Multi-Instance Network (MIN) developed for mammogram classification developed by Lou et. al [2] to lung cancer detection. Their implementation feeds the 6 by 6 by 256 output of the last convolutional layer of AlexNet as instances to three Multi-Instance (MI) loss functions that are then aggregated into a final loss.

We will use the same architecture and first use 2D instances to look for lung cancer. If time permits, we will adapt the MIN to instead use 3D, voxel based convolutions as the output instances for MI loss functions.

We particularly like this AlexNet variant because, amongst the team, we own and have access to two NVidia GTX 970 4GB GPUs.

3.5 Data Augmentation

One major challenge we will face will be the fact that we have a very small collection of data. We plan to heavily augment our data. Inspired by the MIN for mammogram classification, we plan to apply a new random set of augmentation to each epoch of training data.

4 Datasets

The number of lung CT scans available to us is very low. The total number of examples available are many orders of magnitudes smaller than the size of datasets for modern, state of the art classification challenges such as ImageNet or MSCOCO.

Listed below are the datasets we will leverage to train our classification/nodule extraction model.

Dataset	# CT scans	Label Type
Kaggle	1398	Cancer/No-cancer
LIDC-IDRI	1018	Nodule annotation
SPIE	80	Nodule annotation
NLST	?	?

In the dataset above, each CT scan is a stack of approximately 200 2D slices.

The Kaggle dataset will be obtained from the Kaggle Data Science Bowl 2017 competition. The dataset consists of over 1000 CT scans in DICOM format labeled by a cancer/no-cancer binary. We will use this data train our downstream MIN classifier.

The Lung Image Database Consortium (LIDC-IDRI) dataset is the dataset used in the 2016 LUNA challenge.The dataset consists of over 1000 CT scans where each nodule location and radius is annotated for each 2D slice. The dataset was annotated by 4 radiologists and nodules are annotated with varying levels of agreement. We will use this dataset to train our upstream region/nodule proposal network.

The SPIE dataset is a minuscule but have a few important examples where nodules are both annotated and labeled benign or malignant. This dataset may help us validate and visualize the convolutions happening in both the upstream region/nodule proposal and downstream MIN classifier.

We suspect that the National Lung Screening Trial (NLST) dataset will contain about another 1000 CT scans that will either be labeled with the cancer/no-cancer or nodule annotations. We have applied for and are currently waiting for access to the NLST data.

5   Evaluating our results

5.1   Our goals

We have two primary goals in this investigation. First we want to know if region proposal and implicit instance weighting aid multiple instance learning. Second, we want to investigate how leveraging depth information from our data helps with inference with biomedical imaging.

5.2   How will we know what is going on?

The SPIE dataset will give us insight as to how our network sees both benign and malignant nod-

ules or tissues. We hope to visualize both 3D and 2D convolutions in both upstream and downstream networks.

5.3   Quantifying our results

We think that one of the most important contribution this investigation will make is how we can perform biomedical inference with small datasets that are not labeled homogeneously. Our network architecture allows us to plug and play, and even remove, upstream region proposal networks.

We can test different region proposal techniques and quantify how upstream region proposals affect weakly labeled Multi-Instance Networks. We can then also quantify what effects these upstream techniques have on false positive and false negative classifications.

One aspect to focus on specifically is the weight balance on the MI network between the region proposal data and the CT scan data. The proportion of weight allocated to each layer after training leads to some insight about how much the network uses the region proposal information in the final prediction process.

6   Results

6.1   Evaluation of segmention methods

Loss vs Epoch for multiple segmenation networks with LUNA data

Table of segmentation metrics, with IOU and pixel accuracy

Loss vs Epoch for different classifiers

6.2   Visualizations for segmentation

Ground truth vs output visualizations for good cases

Precision vs Recall curves for different classifiers

Ground truth vs output visualizations for bad cases

ROC curves

7   Overall evaluation

Here, we evaluate the effect of upstream ROI proposals. We graph loss over epoch for each of our classifiers with different upstream networks. We also compare these networks with a vanilla classifier without upstream ROI proposals.

We will produce graphs below with Kaggle validation data.

Table of AUC and accuracy

8 About us

8.1 Jay DeStories

Jay is working with Prof. Guyer on graph inference for social marketplaces. His background is primarily in systems and web engineering.

8.2 Jason Fan

The resident math major. Jason is currently working with Prof. Hescott on Graph related kernalized learning. Unfortunately, Jason the only team member who has failed to earn his master's in his fourth year of his undergraduate career. Jason has lead the writing process so far.

8.3 Alex Tong

Alex works with Prof. Hassoun on metabolomics. He is most interested in investigating the effects of the region proposal prior on the performance of the disease prediction network, in relation to models of visual attention [13].

8.4 The Team

In the team, Jay and Jason have already partnered up to investigate a fully convolutional image segmentation technique and will take the lead on adapting region proposal methods to nodule detection. They also own the two GPUs :).

Alex will take the lead in the downstream MIN and investigate other MI loss functions that can enhance cancer detection as well as the effect of region proposal on overall MI network performance.

References

[1] Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, Wieland H. Sommer, Seyed-Ahmad Ahmadi, and Bjoern H. Menze. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3d conditional random fields. *CoRR*, abs/1610.02177, 2016.

[2] Oded Maron and Tomás Lozano-Pérez. A framework for multiple-instance learning. In *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, NIPS '97, pages 570–576, Cambridge, MA, USA, 1998. MIT Press.

[3] M. G. Penedo, M. J. Carreira, A. Mosquera, and D. Cabello. Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. *IEEE Transactions on Medical Imaging*, 17(6):872–880, Dec 1998.

[4] X. Li, L. Shen, and S. Luo. A solitary feature-based lung nodule detection approach for chest x-ray radiographs. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1–1, 2017.

[5] Rushil Anirudh, Jayaraman J. Thiagarajan, Timo Bremer, and Hyojin Kim. Lung nodule detection using 3d convolutional neural networks trained on weakly labeled data, 2016.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[7] Jinsa Kuruvilla and K. Gunavathi. Lung cancer classification using neural networks for {CT} images. *Computer Methods and Programs in Biomedicine*, 113(1):202 – 209, 2014.

[8] Ge Wang He Yang, Hengyong Yu. Deep learning for the classification of lung nodules.

[9] Wentao Zhu, Qi Lou, Yeeleng Scott Vang, and Xiaohui Xie. Deep multi-instance networks with sparse label assignment for whole mammogram classification. *CoRR*, abs/1612.05968, 2016.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CVPR (to appear)*, November 2015.

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

[12] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[13] Volodymyr Mnih, Nicolas Heess, Alex Graves, and koray kavukcuoglu. Recurrent models of visual attention. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and

K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2204–2212. Curran Associates, Inc., 2014.