A (failed) attempt at Lung Cancer Detection with Multi-Instance Networks

Jay DeStories, Jason Fan, Alex Tong

May 12, 2017

1 Part I: The (failed) Attempt

1.1 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. 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 multiple instance learning for whole mammogram classification [1], We investigated two methods for Lung Cancer detection using Multi-Instance Networks.

1.2 Related Work

Over the past decade there has been a significant amount of work towards computer aided diagnosis of lung cancer [2]. 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 [3, 4, 5]. 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 [6, 7].

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 [8].

1.3 Data

-	Unbalanced	Balanced	Test
# Cancer	362	322	50
# No Cancer	1035	322	121
# Total	1397	644	171

Figure 1: Class distributions for datasets

1.4 Multi-Instance Learning

In a malignant lung volume, only a small percentage of the actual tissue is malignant. As such,

it may not be appropriate to learn a cancer/no-cancer binary classification for an entire lung volume. To better model the sparsity of malignant tissue in a lung volume, we instead learn a cancer/no-cancer classification and adjust the loss function appropriately.

We then optimize our model's parameters with the following two loss functions.

1.4.1 Prediction

We will define a model where we deem a lung volume cancerous if there exists a receptive field that is classified to be cancerous. Let F be a set of N feature vectors that encode a receptive field, or an instance, of a given lung slice. Then $\boldsymbol{r}$, the vector of activations, is defined element wise by

$$r_i = \boldsymbol{w}^T \boldsymbol{f}_i + b. \tag{1}$$

Where $\boldsymbol{f}_i$ and r_i is the i-th element of F, and $\boldsymbol{r}$ respectively.

We then use $\boldsymbol{r}$ to predict the probability of cancer $p(y = 1)$ where

$$t = p(y = 1) = \max_{r_i} \ \sigma(r_i). \tag{2}$$

1.4.2 Sparse Binary Cross Entropy Loss

First, we use regular binary cross entropy loss, which is defined by,

$$L_{max} = -y \log(t) - (1 - y) \log(1 - t). \tag{3}$$

Second, we use sparse softmax loss as suggested by Zhu et. al [1], defined by,

$$L_{sparse} = -y \log(t) - (1 - y) \log(1 - t) + \lambda_r \|\boldsymbol{r}\|_1) \tag{4}$$

where the regularization parameter $\lambda_r \|\boldsymbol{r}\|_1$ is scaled by hyperparameter λ_r.

We use this sparse softmax loss because it penalizes a model that classify many instances to be cancerous. Sparse softmax loss encourages the model to only classify few instances to be cancerous which better models our understanding of cancer.

1.5 Method

1.6 Network Architectures

TODO

FC	RF-MIL	Z-MIL
AlexNet Conv5	AlexNet Conv5	AlexNet Conv5
Conv3D (60, 6, 6)	Conv3D (1, 1, 1)	Conv3D (1, 6, 6)
-	Pool (60, 6, 6)	Pool (1,6,6)
Sigmoid	Sigmoid	Sigmoid

Figure 2: Network Architectures

1.6.1 Pre-processing

1.6.2 RGB to Grayscale

When we attempted to use the same technique Zhu et. al. used to extract slice-wise features for lung volumes with pretrained convolutional neural networks (CNN), we could not fit all sixty DICOM slices of a lung volume into VRAM. We ran out of VRAM because, in order to use typical CNNs that expect a 3 channel deep RGB image, the implementation used by Zhu et al. effectively tripled the necessary data volume consumed by the first layer because it naively and redundantly copies each grayscale 224 by 224 slice to each of the channels; However, these redundant copies of the grayscale slices in the RGB channels can be eliminated by manipulating the first convolutional layer of our pretrained CNNs. Let us consider the first convolution applied to one RGB pixel $\boldsymbol{x}$, with weight vector $\boldsymbol{w}$, bias term b, and test time channel-wise mean $\boldsymbol{\mu}$. If all three entries in $\boldsymbol{x}$ have value $\hat{x}$, we can see that,

$$\begin{aligned} & \boldsymbol{w} * (\boldsymbol{x} - \boldsymbol{\mu}) + b \\ &= (\boldsymbol{w} * \boldsymbol{x}) + (b - \boldsymbol{w} * \boldsymbol{\mu}) \\ &= \hat{x} \|\boldsymbol{w}\|_1 + (b - \boldsymbol{w} * \boldsymbol{\mu}). \end{aligned} \tag{5}$$

This means that, with weight vector $\|\boldsymbol{w}\|_1$, bias term $(b - \boldsymbol{w} * \boldsymbol{\mu})$ and no test-time mean, we can convert a redundant 3 channel deep convolution on a grayscale image copied 3 times into the RGB space, to a 1 channel deep convolution! Using this RGB to grayscale conversion, we reduced the VRAM usage in the first layer by a factor of 3 and were able to forward pass entire lung volumes to extract slice-wise features.

1.6.3 Augmentation

1.6.4 Training

1.7 Results

2 Part II: Why did the networks fail?

2.1 HOG features + SVM baseline

2.2 Testing RF-MIL

References

[1] 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.

[2] 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.

[3] 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.

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

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

[6] 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.

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

[8] 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.