LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-682250

Lung Nodule Detection using 3D Convolutional Neural Networks Trained on Weakly Labeled Data

R. Anirudh, J. J. Thiagarajan, P. T. Bremer, H. Kim

February 8, 2016

SPIE Medical Imaging 2016
San Diego, CA, United States
February 27, 2016 through March 2, 2016

Disclaimer

Lung Nodule Detection using 3D Convolutional Neural Networks Trained on Weakly Labeled Data

Rushil Anirudh[1], Jayaraman J. Thiagarajan[2], Timo Bremer[2], and Hyojin Kim[2]

[1]School of Electrical, Computer and Energy Engineering, Arizona State University
[2]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory

ABSTRACT

Early detection of lung nodules is currently the one of the most effective ways to predict and treat lung cancer. As a result, the past decade has seen a lot of focus on computer aided diagnosis (CAD) of lung nodules, whose goal is to efficiently detect, segment lung nodules and classify them as being benign or malignant. Effective detection of such nodules remains a challenge due to their arbitrariness in shape, size and texture. In this paper, we propose a 3D convolutional neural network (CNN) that learns highly discriminative features for nodule detection in lieu of hand-engineered ones such as geometric shape or texture. While 3D CNNs are promising tools to model the spatio-temporal statistics of data, they are limited by their need for detailed 3D labels, which is significantly more expensive than obtaining 2D labels. Existing CAD methods rely on obtaining detailed labels for lung nodules, to train models, which is also unrealistic and time consuming. Therefore, we propose a solution to work with a point label, i.e. the expert needs to specify a single pixel location indicating the presence of the nodule, and its largest expected size. We use unsupervised segmentation to grow out a 3D region, which is used to train the CNN. We show the network trained on the estimated labels have reasonably low false positive rates with a high sensitivity. We demonstrate its performance on the SPIE-LUNGx dataset, which does not have ground truth segment labels.

1. INTRODUCTION

The last decade has seen a significant improvement in using machine learning for computer aided diagnosis (CAD), which can significantly improve efficiency and reduce costs. This has been possible due to the development of feature representations which are work well under several different conditions with invariances to properties such as brightness, shape, and size. More recently, advances in deep neural networks have led to representations that are not hand tuned by an expert but inferred from the training data.[1] These have resulted in significant boosts in accuracy for tasks such as image recognition, natural language understanding, and speech recognition.[1] However, there are significant hurdles before such successes can be transferred to benefit the medical imaging community. A major limiting factor is the difficulty in obtaining annotated data, which is significantly more expensive than compared to traditional computer vision. In this work we address this issue for lung nodule detection, as it is estimated that more people died due to lung and bronchus cancer than all other cancers combined in 2015.[2] Lung nodule detection is also made challenging due to the high variability of nodule shape, size, and texture. As a result, nodule detection techniques that employ classifiers learned using hand-engineered features cannot be easily generalized. Previous methods typically segment out the lung, extract features from the training data and train a classifier to detect potential nodules.[3] More recently some techniques have achieved success using deep learning in one form or the other. For example, Kumar et al.[4] use an autoencoder (an unsupervised learning network) to extract useful features from annotated nodules, these features are used to learn to classify nodules as being malignant or benign. Next, Ginneken et al.[5] have shown promising results using an off-the-shelf convolutional neural network (CNN), one that is pre-trained for an image recognition task. They use the network to obtain features which are used for classification. Two dimensional CNNs have been used in other computer aided detection (CAD) methods such as pancreas segmentation, lymph nodes and colonic polyp detection.[6] Most of these methods are trained individually on 2D images with 2D convolutional filters, whereas the data is inherently 3 dimensional. Roth et. al.[7] have addressed this by considering a '2.5D'

This work was done when the first author was a summer research intern at the Lawrence Livermore National Laboratory in Livermore, CA. Corresponding email: ranirudh@asu.edu

representation that takes slices of the images from a point of interest in 3 orthogonal views. These slices are put together to be treated as a 3-channel image, which is used to train a deep network. In this work, we train a full 3D convolutional network, that can learn 3D convolutional filters from the data. This is beneficial because it can capture the full range of variation expected from the lung nodules. However, there are significant challenges in generalizing 2D convolutional networks to 3D. Firstly, computation is a big hurdle, since 3D convolutions can be expensive, and processing 3D scans (typically $512 \times 512 \times 200$) is significantly more expensive. To circumvent this, we train our network on smaller 3D regions centered around the nodule, we used regions of two different sizes $41 \times 41 \times 7$, and $21 \times 21 \times 7$. Next, a big challenge in medical imaging is the lack of annotations for medical images. This is exaggerated for any 3D processing, since we now require detailed 3D labels which are hard, time consuming and expensive to obtain from experts. Most existing systems for detection require an expert to provide detailed segmentation labels to train a model - which is highly unrealistic and the labels are expensive to obtain. In computer vision, this problem is addressed by outsourcing labeling using services such as Amazons MTurk, which cannot be readily adapted for medical image analysis since experts are required to effectively interpret the data. In order to address this, we have developed a system that reduces the labeling effort on experts by working with "point labels" which are essentially a single pixel location that indicates the presence of a nodule. By using unsupervised learning methods to estimate the *true* label from the weak information, we show that we can reduce the effort required on the expert to label, while being able to train 3D networks that can discriminate effectively. We list our main contributions next.

Contributions:

1. We present a modular system that leverages the robustness of 3D convolutional neural networks for the problem of lung nodule detection. Our system learns the most discriminative features for nodule detection instead of working with hand engineered features such as shape and texture. To the best of our knowledge, we are the first to explore lung nodule detection using 3D convolutional filters.

2. Our system works with point labels, which specify a single voxel location that indicates the presence of a nodule, and its largest cross sectional area. This is much more time efficient compared to the detailed annotations of a nodule in the training set, which is highly impractical since experts are needed to provide these. Using unsupervised learning methods, we estimate a final 3D label which is used to train our 3D CNN.

3. We demonstrate promising results on the AAPM-SPIE-LungX nodule classification dataset.

2. METHODOLOGY

In this section we outline different aspects of our system, that can make predictions on 3D volumes of CT scans. First we address the label estimation procedure for training data. Next we use these estimated labels to train our 3D convolutional network.

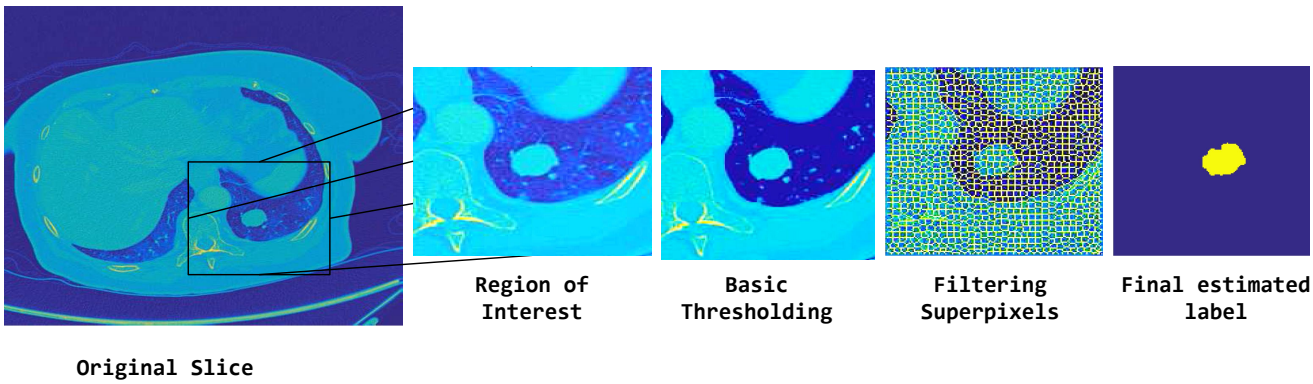


Figure 1: Estimating the ground truth per slice from a point label given by the expert. These 3D labels are used to train the 3D-CNN.

2.1 Estimating weak labels

A limiting factor for using 3D CNNs is the cost of obtaining detailed 3D labels, which are significantly harder to obtain than 2D labels. This is exacerbated by the fact that experts such as radiologists are needed to label lung nodules, as opposed to crowd sourcing platforms such as Amazon MTurk, which have become a norm in computer vision. Instead, we begin by using only a single voxel location or a point label, which indicates the presence of a nodule. We process the slices in $2D$, and combine them using 3D Gaussian filtering. First, we obtain 2D SLIC superpixels[8] to oversegment each slice, as shown in the figure 1. These superpixels find contiguous regions in the image, which are used to eliminate obvious regions that are not nodules based on size and intensity. The 3D Gaussian filtering reduces noise and combines the 2D slices to form a coherent 3D nodule. We are able to do this accurately because we are looking at a local neighborhood around the nodule. The size of the local neighborhood is determined by the largest cross sectional area of the nodule, as given by the expert. The superpixels can effectively aid in capturing nodules that are hard to distinguish at times, such as when they are touching a lung wall.

2.2 Training

The 3D CNN is trained to predict whether or not a single voxel is likely to be a nodule or not, based on the spatio-temporal statistics around it. For example, if the location of the nodule is at $V(x, y, z)$, where V is the entire CT volume, we choose the input volume to be $\hat{v} = V(x - w : x + w, y - w : y + w, z - h : z + h)$, where h is the window size in X, Y planes and h in the Z plane. We used values in the range of $w = 10 - 25$ and $h = 3, 5$. The volume is thinner in the z plane because CT scans are typically sampled much more densely in X, Y planes than in Z. There are at the most 2 nodules per scan, but training a 3D CNN requires many examples to effectively learn the filters. Therefore in order to inflate our training set, we treat different voxels within the same nodule as different positive examples. A typical nodule can range from $3 - 28$ pixels wide at its largest size, and spans $3 - 7$ slices typically, we center our volume at several different randomly sampled voxels within the nodule and pick the resulting volume for a given w, h as a positive training example. Inflating training sets have been useful to train networks that can achieve robustness and avoid overfitting.[7,9] The negative set is much harder to obtain than the positive set because its hard to define it. A negative class contains all examples from the lung that are not the nodule. Fortunately, we can make a few assumptions to make the problem much more tractable. It is important to pick a good negative set containing nodule-like samples, which are likely to end up as false positives. We restrict the negative space to lie within the lung, since it is highly unlikely for a nodule to be found outside it. From within the lung, we random sample locations which have an intensity above a threshold ($\approx 400 - 500$ on the Hounsfield scale). These sampling methods resulted in about $15K$ positive samples and around $20K$ negative samples from the AAPM-SPIE-LungX dataset.[10,11]

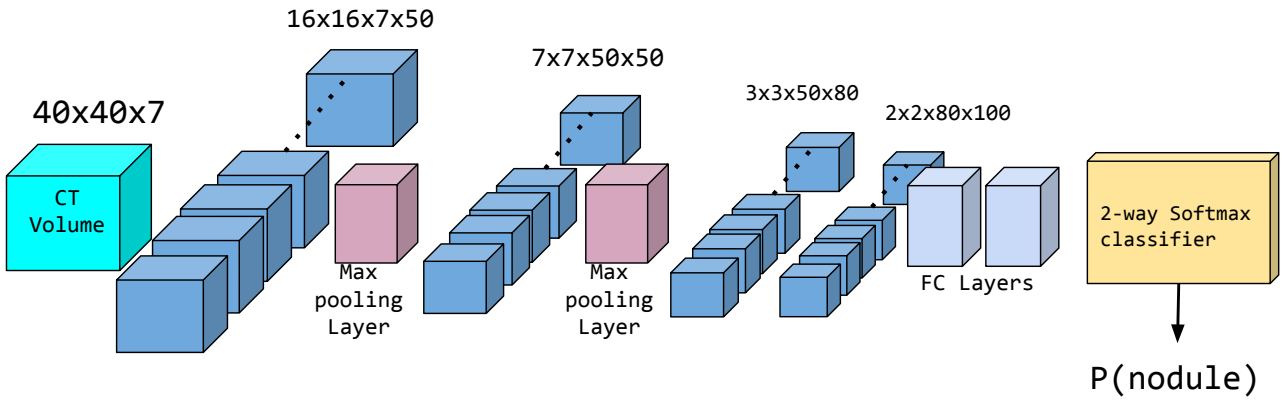


Figure 2: Overall design of the 3D convolutional neural network trained for lung nodule detection.

2.3 Architecture of the Convolutional Neural Net

We trained a 3D CNN using the MatConvNet toolbox for MATLAB.[12] The toolbox allows us to specify the kind of layers, and the number of filters needed. We designed our network to be similar to most of the popular models for image recognition.[9] As shown in figure 2, our network contained 5 convolutional layers which were followed by Rectified Linear Unit (ReLU) activation layers,[9] 2 max-pooling layers, and a final 2-way softmax layer for classification. We also use dropout[13] to regularize the learning problem. Of the five convolutional layers, two are fully connected (FC) with convolution kernels of size 1×1. The generalization from 2D convolutional networks to 3D networks is trivial, in that the filters that are learned are 3 dimensional. Since we use a multiscale approach, we train two different networks for each scale, the convolutional filter sizes for the larger 3D CNN are shown in 2. For the smaller scale, we use the same architecture, and modify the sizes of kernels accordingly.

2.4 Testing and candidate generation

Our network is trained end-to-end, i.e. it is able to make predictions regarding the presence of a nodule directly from a CT volume of the appropriate size. However, the search space in a new test case is extremely large since a typical scan is $512 \times 512 \times 200$ in size. Considering 3D computations are significantly more expensive than 2D, the problem quickly becomes impractical. Instead, we reduce the search space significantly by ruling out parts of the scan that are very unlikely to contain a lung nodule. Since the lung nodule is expected to be inside the lung, we perform lung segmentation using morphological operations on each 2D slice. The segmentation of the lung itself is a hard problem, and there are dedicated systems to perform effective segmentation in 2D and 3D. We also observed that most of the false positive detections on the system were because of the airways which are part of the lung but look a lot like nodules when observed locally. Therefore a robust 3D segmentation can significantly reduce the false positive rate, and improve speed of detection.Next, for each voxel we apply the *dot enhancement filter* using the 3D Hessian. The resulting "dot score" is high if the region around the current voxel is spherical.[6] The dot score map is thresholded in each local neighborhood to provide the final list of candidates. This method can be very effective when the nodules are expected to be approximately round in shape. The dot score is computed as $|\lambda_3|^2/\lambda_1$, where λ_1, λ_2 are the first and second eigenvalues of the 3D Hessian. The dot score for a given volume essentially provides an estimation of its *roundness* such that a high score indicates a tendency towards roundness. We set a low threshold to eliminate obviously non nodule-like elements, and run 3D Gaussian smoothing filter to remove smaller stray particles within the volume. These steps significantly reduce the false positives, resulting in around 80-200 3D nodule like candidates per scan. After smoothing, these can be easily identified using a 3D connected component algorithm efficiently. Finally, we center a test volume at multiple locations inside each candidate and obtain a prediction from the deep network. This also allows us to perform voting in order to eliminate noisy predictions by running a smoothing filter on the predictions.

3. EXPERIMENTS

In this section we describe the dataset, experimental conditions, and results obtained for lung nodule detection.

SPIE-AAPM-LUNGx dataset: The dataset has been published for nodule classification, which requires labeling each nodule as benign or malignant. We use the dataset for detection, as it does not contain detailed labels for nodules, and hence a realistic challenge. Of the 70 scans, we have used 20 for training and 47 for testing. Three scans were discarded because there was ambiguity regarding the presence of a nodule at the specified location. The label is provided as an (x, y, z) location along with information on the largest cross sectional area of the nodule. We did not use this information, however, it could be used to estimate better labels.

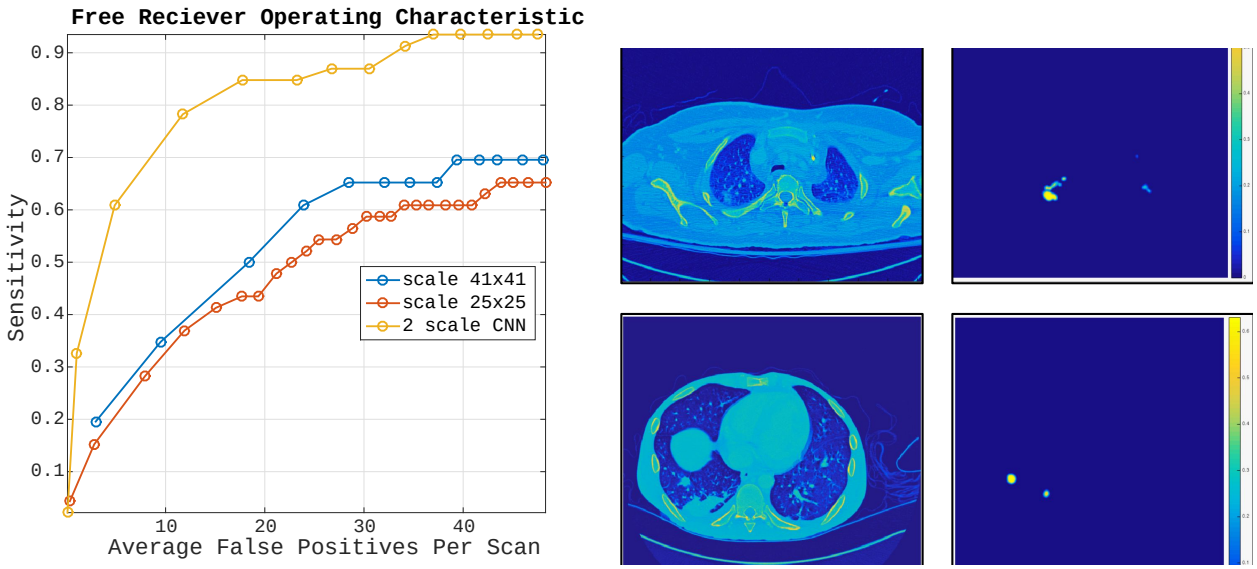
3.1 Evaluation settings

For the test scans, we first generate ground truth labels in a similar fashion as described for the training data. These estimated labels are used to evaluate the performance of our system on the dataset.

Multiscale CNN: The lung nodules vary significantly in size – typically from around 3mm - 20mm. Many successful detection systems employ a multi-scale architecture. Since we are interested in 3D volumes, there are several ways to choose the scale. We chose two scales at $25 \times 25 \times 7$, and $41 \times 41 \times 3$ experimentally. One scale is expected to capture more spatial information while the other, more temporal. We train them separately and

obtain the predictions from each CNN to obtain the final result. The combination performed much better in terms of sensitivity and accuracy as expected. Finally, we generate the free receiver operating curves at various detection thresholds. At a particular threshold, we declare a match if there is a nodule around a small radius (typically $5-10$) of the ground truth. This is done by first estimating the centroid of each 3D blob in the test prediction that is greater than the threshold. Next, we find the distances from each centroid to the ground truth. Only the one that is closest and within a distance threshold is considered a positive, the rest are considered false positives. For each threshold the total number of false positives divided by the number of scans gives the average false positive rate.

Results: We compute the free receiver operating characteristic (FROC) for our system, which plots the sensitivity against the average number of false positives per scan. The results are shown in figure 3a. As it can be seen, even with a weak labeled system, we achieve sensitivities of 80% for 10 false positives per scan. Sample predictions are shown in figure 3.



(a) ROC curve (b) Sample results

Figure 3: Detection performance on the SPIE-AAPM LUNGx dataset.

4. CONCLUSION & FUTURE WORK:

We have presented a system that works with 3D convolutional networks that are trained on point labels, which are significantly easier to obtain. While the initial results look promising, there are areas to further improve the system. Our current system currently processes superpixels in 2D, this could be improved with a 3D superpixel system that clusters coherent spatio-temporal regions. A 3D lung segmentation approach could also eliminate the air tracts which are a primary cause for false positives, but these cannot be differentiated when observed in 2D. Next, the training set can be inflated even further using 3D transforms of existing labels, which has been done for 2D CNNs. Such a technique will also ensure there is little overlap between the original label and transformed label to avoid overfitting.

REFERENCES

[1] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature* **521**(7553), 436–444 (2015).

[2] `http://www.lung.org/lung-disease/lung-cancer/lung-cancer-screening-guidelines/lung-cancer-screening-for-patients.pdf`. [accessed 13-Aug-2015].

[3] Dhara, A. K., Mukhopadhyay, S., and Khandelwal, N., "Computer-aided detection and analysis of pulmonary nodule from ct images: A survey," *IETE Technical Review* **29**(4), 265–275 (2012).

[4] Kumar, D., Wong, A., and Clausi, D. A., "Lung nodule classification using deep features in ct images," in [*Computer and Robot Vision (CRV), 2015 12th Conference on*], 133–138, IEEE (2015).

[5] van Ginneken, B., Setio, A. A., Jacobs, C., and Ciompi, F., "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in [*Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*], 286–289, IEEE (2015).

[6] Choi, W.-J. and Choi, T.-S., "Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor," *Computer methods and programs in biomedicine* **113**(1), 37–54 (2014).

[7] Roth, H. R., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., and Summers, R. M., "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*], 520–527, Springer International Publishing (2014).

[8] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S., "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**(11), 2274–2282 (2012).

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[10] "Spie-aapm-nci lung nodule classification challenge dataset." `https://wiki.cancerimagingarchive.net/display/DOI/SPIE-AAPM-NCI+Lung+Nodule+Classification+Challenge+Dataset`. [accessed 13-Aug-2015].

[11] Armato, III, S. G., Hadjiiski, L., Tourassi, G. D., Drukker, K., Giger, M. L., Li, F., Redmond, G., Farahani, K., Kirby, J. S., and Clarke, L. P., "Guest editorial: Lungx challenge for computerized lung nodule classification: reflections and lessons learned," *Journal of Medical Imaging* **2**(2), 020103 (2015).

[12] Vedaldi, A. and Lenc, K., "Matconvnet-convolutional neural networks for matlab," *arXiv preprint arXiv:1412.4564* (2014).

[13] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research* **15**(1), 1929–1958 (2014).

Prepared by LLNL under Contract DE-AC52-07NA27344.