Lung Nodule Detection based on Convolutional Neural Networks

Julio Mendoza Bobadilla
Advisor: Prof. Helio Pedrini

Master’s Defense
Institute of Computing
University of Campinas

December 18, 2017
Summary

1. Introduction
   • Problem
   • Motivation
   • Hypotheses
   • Contributions

2. Background
   • CAD Systems
   • Lung Area Segmentation
   • Candidate Nodule Localization
   • Candidate Nodule Classification

3. Methodology

4. Results

5. Conclusions
   • Future Work
Introduction

Problem

Input: Chest X-Ray

Output: Potential nodules highlighted
Estimated number of cancer deaths on both sexes (Source: World Cancer Report 2014\textsuperscript{1}).

- Improve the performance of radiologists on lung cancer screening.
- Address the drawbacks of lung cancer screening with CT images.
- Explore the potential role of CAD systems with CXR for lung cancer screening.
- Leverage the advantages of deep learning approaches on lung nodule detection.

Introduction
Hypotheses

Main hypothesis:

Convolutional Neural Networks (CNNs) trained from the scratch can perform well on lung nodule classification by setting proper regularization and optimization methods.

Secondary hypotheses:

- The design of a specialized architecture for the classification stage of lung nodule detection improve the performance of our method.
- The regularization effects of data augmentation, dropout and weight penalties critical for lung nodule classification due to the amount of samples used in training.
- The regularization and optimization effects of unsupervised objectives in loss functions improve CNN performance in lung nodule classification.
Introduction
Contributions

Main contributions:

- An analysis and evaluation of methods for lung area segmentation and candidate nodule localization.
- The proposition of a method for lung nodule classification based a CNN trained from the scratch.
- The comparison of the our lung nodule detection pipeline with other approaches of the literature.
Usage modalities:
- As first reader: radiologists analyze the regions detected by the CAD system.
- As second reader: CAD system may detect nodules missed by radiologists.

CAD systems usually detect nodules solving three subproblems:
Background
Lung Area Segmentation

Lung Area Segmentation → Candidate Nodule Localization → Candidate Nodule Classification
Background

Active Appearance Model

- **Shape model**
  - Normalization with Procrustes Analysis.
  - Finding bases with Principal Component Analysis.

  \[
  s = \bar{s} + \sum_{i=1}^{n} p_i s_i
  \]

  - Mean shape
  - Shape bases

- **Appearance model**
  - Appearance representation using visual descriptors.
  - Finding bases with Principal Component Analysis.

  \[
  A = \bar{A} + \sum_{i=1}^{m} c_i A_i
  \]

  - Mean appearance
  - Appearance bases

- **Model instantiation**

  \[
  M(W(x, p)) \simeq \bar{A} + \sum_{i=1}^{m} c_i A_i
  \]

- **Model fitting**

  \[
  p^*, c^* = \arg \min_{p,c} \left\| M(W(x, p)) - \left( \bar{A} + \sum_{i=1}^{m} c_i A_i \right) \right\|^2
  \]
Background

Active Appearance Model

\[ \text{Appearance, } A = A_0 + 3559A_1 + 351A_2 - 256A_3 \ldots \]

\[ \text{Shape, } s = s_0 - 54s_1 + 10s_2 - 9.1s_3 \ldots \]

Source: Matthews and Baker\(^2\)

Background
Candidate Nodule Localization

Lung Area Segmentation
Candidate Nodule Localization
Candidate Nodule Classification
Background
Scale-Space based Detectors

\[ L(\cdot, \cdot; t) = g(\cdot, \cdot; t) \ast f(\cdot, \cdot) \]

Laplacian of Gaussian

Scale-space blob detection.

- normalized Laplacian of Gaussian (LoG): \( t \nabla L = t(L_{xx} + L_{yy}) \)
- normalized Determinant of Hessian (DoH): \( t \nabla L = t^2(L_{xx}L_{yy} - L_{xy}^2) \)
- Difference of Gaussians (DoG): \( DoG(t) = L(\cdot, \cdot; t + \Delta t) - L(\cdot, \cdot; t) \)
Background
Convergence Index based detectors

- Convergence degree at point $Q$: $\cos \theta(k, l)$
- Convergence index at point $P$ with support region $R$:

$$CI(i, j) = \frac{1}{M} \sum_{(k,l) \in R} \cos \theta(k, l)$$
Convergence index at point $P$ with support region $R$:

$$SB(x, y) = \frac{1}{N} \sum_{i=0}^{N-1} \max_{r_{\text{min}} \leq n \leq r_{\text{max}}} \left( \frac{1}{d} \sum_{m=n}^{n+d} \cos \theta_{i,m} \right)$$

CI of a band in the line $B_i$
Background
Candidate Nodule Classification

Lung Area Segmentation → Candidate Nodule Localization → Candidate Nodule Classification
Feature extraction: geometric, shape, intensity and gradient features.

Binary classification.
Background
Feature Engineering

Input nodules → Nodule Segmentation → Feature Extraction → Feature Selection / Reduction → Classifier
Background
Convolutional Neural Networks

- Convolutional Neural Networks
  - Convolution
  - Pooling

Input nodules → ConvNet → ...
Methodology
Overview

- We segment the lung area using patch-based AAM.
- We find candidate locations with an SB filter.
- We estimate the probability of nodules being candidates using a CNN.

Overview of the main steps of the proposed methodology.
Methodology
Segmentation

Training

Images
Landmarks

Multi-Scale Patch-based Active Appearance Model

Modeling shape and appearance

AAM parameters
\[
\bar{s} \in \mathbb{R}^{2n \times 1}, \quad \bar{a} \in \mathbb{R}^{F \times 1}, \\
S \in \mathbb{R}^{2n \times n}, \quad A \in \mathbb{R}^{F \times n}
\]

Testing

Input

Wiberg optimization
\[
\arg\min_{p,c} \| i[p] - (\bar{a} + Ac) \|^2
\]

Fitting model

Landmarks

Masks

Overview of the segmentation stage.
Visualization of Patch-based AAM (for low and high resolution models) fitted on the sample JPCLN001.IMG from the JSRT dataset.
Methodology

Detection

Input X-ray (JPCLN001.IMG)  SB filter output

Candidate nodules detected with an SB filter.

Non-maxima suppression with $r_{sbf} = 7\text{px}$
Overview of the classification stage.

Training

Candidate locations \(\{(x, y)\}\)
- Images
- Nodule locations

Candidate ROIs
- Labeling
- Data augmentation
- Balancing samples
- Training CNN
- ConvNet

Testing

Input
- Segmentation and Candidate Detection
- Candidate ROIs
- ConvNet
- Top 4 candidates

Candidates with probs \(\{(x, y, \text{prob})\}\)
Methodology
Data Preparation and Augmentation

Data preparation:
- Labeling criterion: 25mm.
- Z-score normalization.

Data augmentation:
- Affine transformations (translation, rotation, scaling, flip).
- Intensity shifts.
Methodology
Architecture Design

The network is composed of $c$ alternating convolutional (in red), and max-pooling (in green) layers, followed by $f$ fully connected layers (in yellow).

- The network is fed with images with size $2^{c+1} \times 2^{c+1}$. 

Representation of the architecture $ConvNet(c, k, f)$. 
Methodology

Learning

Algorithm 1 Stochastic Gradient Descent with balanced samples.

1: procedure SGD-BALANCEDSAMPLES
2: Require: Network parameters $\theta$
3: Require: Positive samples $P$
4: Require: Negative samples $N$
5: while AUC does not stop improving do
6:  Select the first half of the minibatch from $N$ iteratively.
7:  Select the second half of the minibatch from $P$ randomly.
8:  Apply the data augmentation transformations to all samples of the minibatch.
9:  Compute gradient estimate with the minibatch.
10: Update network parameters $\theta$.
11: end while
12: end procedure
Composed of 247 CXR images. 154 images with nodules and 93 without nodules.

The annotations include nodule location, level of subtlety, effective diameter, malignancy condition, among others.

Patches with nodules extracted from CXRs of the JSRT dataset.
Results

SCR Dataset

Left: The initial points marked by an observer on the first image of the JSRT database. Right: the interpolated landmarks along the contour for use of segmentation methods (Source: Van Ginneken, Stegmann, and Loog\(^3\)).

- Created to evaluate method for segmentation of lungs, hearth, and clavicles in CXR images.
- Composed of landmarks extracted from the images of the JSRT dataset.

Results

LIDC-IDRI Dataset

- Contains 1018 CT scans, and 290 CXR associated with some of these CT scans.
- The annotations include nodule location, level of subtlety, effective diameter, malignancy condition, annotation confidence, among others.
- We filtered the annotation:
  - Effective diameter $\geq 3\text{mm}$
  - Accepted by at least 3 radiologists.
- We included nodules annotated by 2 radiologists when the sum of the confidence level $\geq 7$
- We excluded nodules that have an average subtlety level $< 2$.  

Patches with nodules extracted from CXRs of the LIDC-IDRI dataset.
Segmentation results on the positive subset of JSRT dataset.

Segmentation results on the LIDC-IDRI dataset.
### Results

#### Segmentation

<table>
<thead>
<tr>
<th>Method</th>
<th>Ω</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Shape</strong></td>
<td>0.714 ± 0.075</td>
</tr>
<tr>
<td><strong>ASM tuned</strong></td>
<td>0.927 ± 0.032</td>
</tr>
<tr>
<td><strong>Pixel Classification Post-Processed</strong></td>
<td>0.945 ± 0.022</td>
</tr>
<tr>
<td><strong>Human Observer</strong></td>
<td>0.946 ± 0.018</td>
</tr>
<tr>
<td><strong>Non-Rigid Registration using Atlas</strong></td>
<td>0.954 ± 0.015</td>
</tr>
<tr>
<td><strong>Level Set + Deep Belief Network</strong></td>
<td>0.985 ± 0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>μ ± σ</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proposed method (Patch-based AAM with DSIFT)</strong></td>
<td>0.935 ± 0.019</td>
<td>0.786</td>
<td>0.964</td>
</tr>
<tr>
<td><strong>ASM tuned</strong></td>
<td>0.927 ± 0.032</td>
<td>0.745</td>
<td>0.964</td>
</tr>
<tr>
<td><strong>Human Observer</strong></td>
<td>0.946 ± 0.018</td>
<td>0.822</td>
<td>0.972</td>
</tr>
<tr>
<td><strong>Non-Rigid Registration using Atlas</strong></td>
<td>0.954 ± 0.015</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Level Set + Deep Belief Network</strong></td>
<td>0.985 ± 0.003</td>
<td>0.972</td>
<td>0.991</td>
</tr>
</tbody>
</table>

Segmentation results in terms of the Jaccard coefficient Ω. Methods are ranked according to their mean.

---


FROC curves for detectors evaluated using LIDC-IDRI dataset.
We evaluate the performance of various data augmentation schemes based on including transformations incrementally.
Results
Classification: Dropout Schemes

Comparing constant dropout with linear increasing dropout on convolution layers.
Results
Classification: Architecture Design

FROC curves varying the parameters $c$, $k$ and $f$. 

- Varying $c$ with $k = 64$ and $f = 1$
- Varying $k$ with $c = 6$ and $f = 1$
- Varying $f$ with $c = 6$ and $k = 32$
Results
Classification: Unsupervised Objectives

Experiments considering supervised and unsupervised objectives on small CNN (ConvNet(3, 32, 2)).
## Results

### Classification: Sources of Variation on Evaluation Protocols

<table>
<thead>
<tr>
<th>CAD System</th>
<th>Datasets</th>
<th>Protocol</th>
<th>Labeling Criteria</th>
<th>Parameter Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen and Suzuki(^7)</td>
<td>Private dataset, JSRT excluding 14 opaque cases</td>
<td>Train on private dataset, test on JSRT subset</td>
<td>Distance (25 mm)</td>
<td>Considers test results</td>
</tr>
<tr>
<td>Hardie et al.(^8)</td>
<td>Private dataset, JSRT excluding 14 opaque cases</td>
<td>Train on a private dataset, test on JSRT, 10-fold cross-validation on JSRT subset</td>
<td>Distance (25 mm)</td>
<td>Considers only training set</td>
</tr>
<tr>
<td>Schilham et al.(^9)</td>
<td>JSRT dataset</td>
<td>5-fold cross-validation on JSRT</td>
<td>Overlap (&gt; 0%)</td>
<td>Considers test results</td>
</tr>
<tr>
<td>Shiraishi et al.(^10)</td>
<td>Private dataset augmented with JSRT images</td>
<td>Train and test on the merged dataset</td>
<td>Distance (22mm and 24mm)</td>
<td>Considers only training set</td>
</tr>
<tr>
<td>Coppini et al.(^11)</td>
<td>Private dataset, JSRT subset of 140 samples</td>
<td>5 partitions of the JSRT subset</td>
<td>Centroid</td>
<td>Considers test results</td>
</tr>
<tr>
<td>Wei et al.(^12)</td>
<td>JSRT dataset</td>
<td>Leave-one-out cross-validation</td>
<td>Unspecified</td>
<td>Considers test results</td>
</tr>
<tr>
<td>Wang et al.(^13)</td>
<td>JSRT excluding opaque cases</td>
<td>10-fold cross-validation</td>
<td>Distance (unspecified)</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

Main sources of variation of the evaluation protocol used in competing methods.

## Results

### Classification Considering Only Supervised Objectives

CAD system performance comparison.

<table>
<thead>
<tr>
<th>Average FPPI</th>
<th>Method</th>
<th>Reported Sensitivity (%)</th>
<th>CNN Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>Chen and Suzuki(^a,c)</td>
<td>85.0</td>
<td>90.0</td>
</tr>
<tr>
<td>5.0</td>
<td>Hardie et al.(^a)</td>
<td>80.1</td>
<td>90.0</td>
</tr>
<tr>
<td>2.0</td>
<td>Schilham et al.(^b)</td>
<td>51.0</td>
<td>71.4</td>
</tr>
<tr>
<td>4.0</td>
<td>Schilham et al.(^b)</td>
<td>71.0</td>
<td>79.2</td>
</tr>
<tr>
<td>5.0</td>
<td>Shiraishi et al.(^a,b)</td>
<td>70.1</td>
<td>90.0</td>
</tr>
<tr>
<td>4.3</td>
<td>Coppini et al.(^c)</td>
<td>60.0</td>
<td>79.9</td>
</tr>
<tr>
<td>5.4</td>
<td>Wei et al.(^d)</td>
<td>80.0</td>
<td>82.5</td>
</tr>
<tr>
<td>1.19</td>
<td>Wang et al.(^e)</td>
<td>69.3</td>
<td>72.1</td>
</tr>
</tbody>
</table>

\(^a\) Results reported in this row exclude opaque cases.

\(^b\) Results based on 924 chest radiographs that include the JRST cases.

\(^c\) A private database was used for training and the JSRT for testing.

---

Comparison of the method performance on the JRST database. Sensitivity values are adjusted by considering opaque cases as missed.
Results
Classification

Visual results of our system. We highlight the 4 top positive candidates detected by each image. The saturation of each bounding box is proportional to the probability of the candidate of being a true nodule.
Conclusions

Final Remarks

- Our method aimed to detect potential location of nodules with high sensitivity through a few false positives per image.
  - We analyzed and compared methods for lung area segmentation and candidate nodule classification.
  - We explored the effectiveness of CNNs for reducing false positives on the candidate classification stage.
- We showed that CNN trained from the scratch obtained good results on lung nodule classification.
  - We analyzed the impact of various data augmentation transformations.
  - We compared two ways to use dropout on convolutional layers. We found that assigning increasing dropout probabilities converged faster and it was slightly superior than assigning fixed dropout.
  - We evaluated the impact of CNN architecture parameters. We obtained best results with ConvNet(6, 32, 1).
  - We trained all models balancing batches on each SGD iteration.
  - We showed that augmenting the loss function with an unsupervised term improved the effectiveness of the methods.
- We compared the results of the best configuration with competing approaches.
  - A fair comparison is difficult due to variation on evaluation protocols of the methods.
  - Under our considerations with respect to variations, we showed that our method achieved good results when compared to the state-of-the-art.
Conclusions

Future Work

- Results obtained by considering supervised and unsupervised objectives on CNN training suggested that further investigation in this direction should be conducted.
- Results found in this dissertation can be leveraged to improve methods for lung nodule classification in CT images.
Acknowledgements

- IC/UNICAMP for the infrastructure.
- CAPES for the financial support.