Unsupervised Discovery of Novel Emphysema Subtypes

Andrew Francis Laine, D.Sc.
Percy K. and Vida L. W. Hudson Professor of Biomedical Engineering and
Professor of Radiology (Physics)
Department of Biomedical Engineering
Columbia University
New York, NY 10027

Presented at
New Jersey Public Health Annual Conference:
“Pressing Public Health Challenges: Systems Thinking, Opioid Crisis, and Use of Artificial Intelligence”
Rutgers Continuing Education Center at Atrium, Division of Continuing Education
October 4, 2019, Somerset, NJ
Weakly-Supervised Lung Texture Learning to Localize Nodules and Detect Lung Cancer

Challenges

- Weak-labels are less informative than voxel-level delineation.
- False positive of nodule detection due to high attenuation tissues.

Contributions

- Proposed a weakly-supervised framework for lung nodule detection, and achieved competitive performance compared to a fully-supervised method.
- Proposed effective lung cancer prediction approaches at scan-level and nodule-level.

References:

Method
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

1 Nodule activation maps (NAMs) [1]

\[ A_k = \sum_{(x,y)} a_k(x,y) \]

\[ S_{\text{nodule}} = \sum_k w_{k,\text{nodule}} \cdot A_k \]

\[ \text{NAM}(x,y) = \sum_k w_{k,\text{nodule}} \cdot a_k(x,y) \]

Method
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

1. Nodule activation maps (NAMs)

[Diagram showing the process of using a Multi-GAP CNN to detect nodules in lung CT images.]
Method
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

1. Nodule activation maps (NAMs)

Discriminative detection
Accurate localization
Method
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

2 Nodule candidate screening

- Test image
- Nodule
- Coarse segmentation
- Candidate Masking
- Ground truth vs. Fine segmentation
- R-NAM

CNN
Method

Weakly-supervised lung texture learning to localize nodules and detect lung cancer

Weakly-supervised CNN:
- Based on VGG-16

Network Architectures

Fully-supervised CNN:
- Based on U-net, widely used for semantic segmentation.
Method
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

NAM-based Individual-Level Nodule-Specific Lung CT Signature

- Interpretable & low-dim representations of the lung CT scans;
- Suited to work with scan-level diagnostic labels:
  - Kaggle Data Science Bowl 2017: can you predict lung cancer?

Potential Nodule:
Large, locates at superior, anterior and peel region of the right lung.
Experimental Results
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

Data - LIDC-IDRI dataset:

- 1,010 thoracic CT scans:
  - Voxel-level delineation of nodules;
  - Can generate weak-labels.
- 8,345 slices with nodule;
- 8,345 slices no nodule;

\[ \text{TPR} = \text{true positive rate of nodule detection}; \]
\[ \text{FPR} = \text{false positive rate on slice without nodules}; \]
\[ \text{FPR}_\text{nodule} = \text{false positive rate on slices with nodule}; \]
\[ \text{TP Dice} = \text{Dice on truly detected nodules}; \]
\[ \text{TP DOA} = \text{Difference of area on truly detected nodules}. \]

* = best performance with our framework;

**boldfaced** = overall best performance;

<table>
<thead>
<tr>
<th>Method</th>
<th>TPR</th>
<th>FPR</th>
<th>FPR_nodule</th>
<th>Dice mean ± SD</th>
<th>TP Dice mean ± SD</th>
<th>TP DOA mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-GAP Coarse</td>
<td>0.77*</td>
<td>0.11*</td>
<td></td>
<td>0.46 (±0.31)</td>
<td>0.61 (±0.20)</td>
<td>57.6 (±71.1)</td>
</tr>
<tr>
<td>2-GAP Coarse</td>
<td>0.76</td>
<td>-</td>
<td></td>
<td>0.50 (±0.34)</td>
<td>0.66 (±0.18)</td>
<td>41.6 (±53.6)</td>
</tr>
<tr>
<td>3-GAP Coarse</td>
<td>0.75</td>
<td>-</td>
<td></td>
<td>0.50 (±0.32)</td>
<td>0.67 (±0.18)</td>
<td>40.1 (±50.9)</td>
</tr>
<tr>
<td>1-GAP Fine</td>
<td>0.75</td>
<td>-</td>
<td>0.14*</td>
<td>0.54 (±0.34)</td>
<td>0.73 (±0.15)</td>
<td>30.7 (±52.8)</td>
</tr>
<tr>
<td>2-GAP Fine</td>
<td>0.75</td>
<td>-</td>
<td>0.14</td>
<td>0.55* (±0.33)</td>
<td>0.74* (±0.14)</td>
<td>29.2* (±46.8)</td>
</tr>
<tr>
<td>3-GAP Fine</td>
<td>0.74</td>
<td>-</td>
<td>0.15</td>
<td>0.54 (±0.34)</td>
<td>0.74 (±0.14)</td>
<td>29.3 (±46.4)</td>
</tr>
<tr>
<td>U-net</td>
<td>0.74</td>
<td>0.29</td>
<td>0.26</td>
<td><strong>0.56 (±0.38)</strong></td>
<td><strong>0.76 (±0.19)</strong></td>
<td><strong>28.3 (±44.8)</strong></td>
</tr>
</tbody>
</table>

Better performance on nodule detection

Competitive performance on nodule segmentation
Experimental Results
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

Data - Kaggle DSB2017 dataset:
- 2,101 thoracic CT scans:
  - 1,397 training, 198 validation (stage 1) and 506 test (stage2)
- Ground truth = cancer diagnosis.
  - One binary label per scan;
  - With/without early lung cancer.

Results:
- Log loss = 0.45872
- Rank in Kaggle Data Science Bowl 2017: 14th out of 1,972 teams

Additional annotations of “mass”.
Nodule-level detailed characterization.
Individual-level spatial activation is beneficial.
Experimental Results
Weakly-supervised lung texture learning to localize nodules and detect lung cancer

Visualization of CNN filters learned in the VGG-16 model:
- Visualized via activation maximization
**Aim:**
- Predict malignancy of detected nodules in CT images.

**Challenge:**
- Small number of training data, especially for malignant cases.

**Method:**
- Transfer learning with 3D classification CNN;
- Class-aware nodule inpainting.
Characterizing Early Lung Cancer at Nodule-Level
* Side Project with Siemens Corporate Research
Quantitative Evaluation on LIDC dataset
• 1791 nodules:
  • 1506 benign vs. 285 malignant
• Training vs. validation vs. test = 4:1:1
• Add synthetic malignant nodule patches

<table>
<thead>
<tr>
<th>Dataset Splits</th>
<th>Benign</th>
<th>Malignant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>1,004</td>
<td>191</td>
<td>1,195</td>
</tr>
<tr>
<td>Train+Syn</td>
<td>1,004</td>
<td>191 + 463</td>
<td>1,658</td>
</tr>
<tr>
<td>Validation</td>
<td>251</td>
<td>47</td>
<td>298</td>
</tr>
<tr>
<td>Test</td>
<td>251</td>
<td>47</td>
<td>298</td>
</tr>
</tbody>
</table>

Classification Results

<table>
<thead>
<tr>
<th>Network</th>
<th>ACC</th>
<th>SEN</th>
<th>SPE</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.859</td>
<td>0.660</td>
<td>0.896</td>
<td>0.862</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>0.861</td>
<td>0.653</td>
<td>0.901</td>
<td>0.847</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>0.873</td>
<td>0.596</td>
<td>0.924</td>
<td>0.860</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>ACC</th>
<th>SEN</th>
<th>SPE</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Training + Weighted Loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.842</td>
<td>0.681</td>
<td>0.873</td>
<td>0.836</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>0.826</td>
<td>0.723</td>
<td>0.847</td>
<td>0.810</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>0.829</td>
<td>0.702</td>
<td>0.853</td>
<td>0.818</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>ACC</th>
<th>SEN</th>
<th>SPE</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Training + Synthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.883</td>
<td>0.702</td>
<td>0.916</td>
<td>0.867</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>0.893</td>
<td>0.702</td>
<td>0.928</td>
<td>0.881</td>
</tr>
<tr>
<td>ResNet-152</td>
<td><strong>0.903</strong></td>
<td>0.660</td>
<td><strong>0.948</strong></td>
<td><strong>0.883</strong></td>
</tr>
</tbody>
</table>
Summary - Contributions and Impacts

• Unsupervised learning of localized emphysema texture patterns:
  • Novel lung shape spatial mapping = a useful tool to study spatial patterns on lung CT.
  • Novel discovery of 10 highly-reproducible sLTPs and 6 clinically-significant QES:
    • May facilitate disease understanding and personalized therapy.

• Labeling emphysema texture on cardiac CT scans:
  • Robust emphysema segmentation on cardiac CT scans;
  • Novel lung texture labeling with domain adaptation on cardiac CT scans:
    • Enable usage of widely available cardiac CT scans.

• Novel weakly-supervised lung nodule segmentation:
  • Transforming ML field with less annotation when training on evolving scanner technologies.
  • Practical validation of predicting lung cancer at individual-level and nodule-levels.
References


Acknowledgements

Heffner Biomedical Imaging Lab
Elsa D. Angelini Ph.D.
Xinyang Feng (Ph.D. student)
Yrjö Häme (Ph.D. student)
Jingkuan Song (Post-doc)
Yu Gan (Post-doc)
Thomas Vetterli (M.Sc. Intern)

Collaborators

Columbia University Medical Center:
  R. Graham Barr, M.D., Ph.D.
  Pallavi P. Balte, Ph.D.
  John H.M. Austin, M.D.
  Benjamin M. Smith, M.D.
  Yifei Sun, Ph.D.
  Wei Shen, M.D.

Iowa University:
  Eric A. Hoffman, Ph.D.
  Mark Escher, B.A.

University of Virginia:
  Ani Manichaikul, Ph.D.

And the MESA, SPIROMICS Investigators!
Extra Slides
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

• Learn sLTPs in MESA COPD study:
  • N = 317 full-lung CT scan

12 sLTPs discovered, ordered by average intensity

• Training ROI: %\textit{emph} > 1%

Columbia University
In the City of New York
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

• Learning reproducibility:
  • $R_{ln}$ among training subsets (50% scans overlap):
    • Texture only: 0.85
    • Texture + Spatial: 0.78

• Labeling reproducibility:
  • $R_{la}^{DC}$ = average Dice of sLTP masks over all test scans;
  • $R_{la}^{CC}$ = Spearman’s correlation of %sLTP over all test scans.

$$R_{ln} = \frac{1}{N_{set} \cdot N_{sLTP}} \sum_{c=1}^{N_{set}} \sum_{k=1}^{N_{sLTP}} \frac{|\Lambda_{sLTP_k} \cap \Lambda_{\pi(sLTP^c_k)}|}{|\Lambda_{sLTP_k}|}$$
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

- Association of sLTPs with standard subtypes sLTPs in MESA COPD study:
  - N = 317 full-lung CT scans.

\[ \text{Compared to:} \]
- Initial 100 LTPs;
- Our previous work using texture + spatial co-occurrence.
  - Method A: Hame, et al. 2015;
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

Global label

\[ H_g = [\%CLE, \%PLE, \%PSE, \%NE] = [P(L(x) = C_1), ..., P(L(x) = C_4)] \]

sLTP histogram

\[ H_p = [\%sLTP_1, \ldots, \%sLTP_{12}, \%NE] = [P(F(x) = p_1), \ldots, P(F(x) = p_{13})] \]

\[ P(L(x) = C_i) = \sum_{k=1}^{13} P(L(x) = C_i | P(F(x) = p_k))P(F(x) = p_k) \]

\[ H_g = H_p A \]

\[ A_{k, i} = P(L(x) = C_i | F(x) = p_k) \]
Unsupervised Learning

- GWAS results:
  - 5 genetic variants for four QES
  - Apical QES: DRD1
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

- ADC vs. QES labeling
- Enlarged alveolar size
Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema
HMMF-based Emphysema Segmentation

Bayes rule:

\[ P(f | I) = \frac{P(I | f)P(f)}{P(I)} \]

Likelihood

Prior

\[ P(q, \theta | I) = \frac{1}{R} P(I | q, \theta) P(q) P(\theta) \]

\[ P(q) \propto -\lambda K(r_1, r_2) \sum_{k=1}^{\mathcal{C}} (q_k(r_1) - q_k(r_2))^2 \]

- \( \mathcal{C} = \text{clique} \) = 8-connected neighborhoods in a 2-D plane with slice thickness ~2.8mm.
- \( \lambda = \text{Markovian weight} \) which depends on image quality and noise level and needs to be scanner-specific.

Key ideas:

- Define 3 scanner categories:
  - \( S_B \) = used only at Baseline;
  - \( S_{BF} \) = used at Baseline-Follow-up;
  - \( S_F \) = used only at Follow-up.

- Hypothesis: normals at baseline have population average %emph \( m_B = 2\% \)

- Empirically adjust \( \lambda_B \) to match target value \( m_B(\lambda_B) = 2\% \)
- Transfer in time: \( \lambda_{BF} = \lambda_B \)
- Equalize in time: \( m_{BF}(\lambda_B) = m_F(\lambda_F) \)
- Use temporal prediction for FL scans \( m_{FL}(\lambda_F) = m_{\text{predict}} \)

\( z = \text{constant} \)
\( K = \text{radial kernel} \)
Weakly-supervised lung texture learning to detect nodules and lung cancer

1-GAP Detection Performance over Network Architectures

- **DenseNet-121:**
  - Easier to train but less discriminative.
- **ResNet-50:**
  - Hard to converge; higher TPR but also higher FPR.
- **VGG-16:**
  - Overall best detection performance.

<table>
<thead>
<tr>
<th>Method</th>
<th>Validation ACC</th>
<th>Test ACC</th>
<th>TPR</th>
<th>FPR</th>
<th># of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-16</td>
<td>0.887</td>
<td>0.884</td>
<td>0.77</td>
<td>0.11</td>
<td>15,242,050</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.891</td>
<td>0.870</td>
<td>0.78</td>
<td>0.17</td>
<td>25,687,938</td>
</tr>
<tr>
<td>DenseNet-121</td>
<td>0.892</td>
<td>0.873</td>
<td>0.74</td>
<td>0.13</td>
<td>8,089,154</td>
</tr>
</tbody>
</table>
PDCM versus Locations of Nodule Types

LIDC-IDRI dataset:
- 1791 nodules:
  - 1506 benign
  - 285 malignant

Malignant nodules: less in the external border

Malignant nodules: more in superior lung

No significant difference
Background Concepts
Texton-based Features

Texton codebook

ROI

Texton-based feature
\[ h^l_i = \sigma \cdot \left( \sum_k h^{l-1}_k \ast W^l_{ki} + b^l_i \right) \]

\[ h^{l-1}_i = \text{i-th feature map in layer } l \]

\[ h^{l-1}_k = \text{k-th feature map in layer } l - 1 \]

\[ h^l = \sigma \cdot \left( \sum_k h^{l-1} W^l + b^l \right) \]

\[ h^{l-1} = \text{feature vector } \in \mathbb{R}^P \text{ in layer } l-1 \]

\[ h^l = \text{feature vector } \in \mathbb{R}^Q \text{ in layer } l \]