Exploiting Visual and Report-Based Information for Chest X-Ray Analysis by Jointly Learning Visual Classifiers and Topic Models

Zachary A. Daniels, Dimitris N. Metaxas
zad7@cs.rutgers.edu, dnm@cs.rutgers.edu

Center for Computational Biomedicine Imaging and Modeling
Department of Computer Science
Rutgers, The State University of New Jersey

Motivation and Problem Overview

- Visual inspection of chest x-rays is a common and important method for diagnosing certain life-threatening diseases such as pneumonia.
- Manual examination of chest x-rays is time-consuming and requires significant effort by highly-trained radiologists.
- (Semi-)automated chest x-ray analysis using computer vision and machine learning can act as a support tool for radiologists.
- Many chest x-rays come with radiologist-dictated reports that justify the diagnostic impressions.
- We propose a new method that exploits both visual and textual information for improved automated chest x-ray analysis.
- We propose a DNN architecture that simultaneously:
  1. **Constructs a topic model** which clusters key terms from the findings into meaningful groups
  2. **Predicts the presence or absence of each topic** for a given image based on learned visual features
  3. Uses an image's predicted topic encoding as features for **predicting one or more diagnoses**
- Since the net learns the topic model jointly with the classifier, it gives us a powerful tool for understanding which semantic concepts the net might be exploiting when making diagnoses.
- Since the net is constrained to predict topics based on expert-annotated reports, it automatically encodes some higher-level expert knowledge about how to make diagnoses.

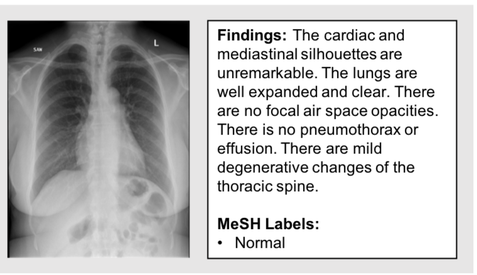


Fig: An example chest x-ray paired with a radiologist's natural language findings and class labels

Extracting Key Terms from Natural Language Text

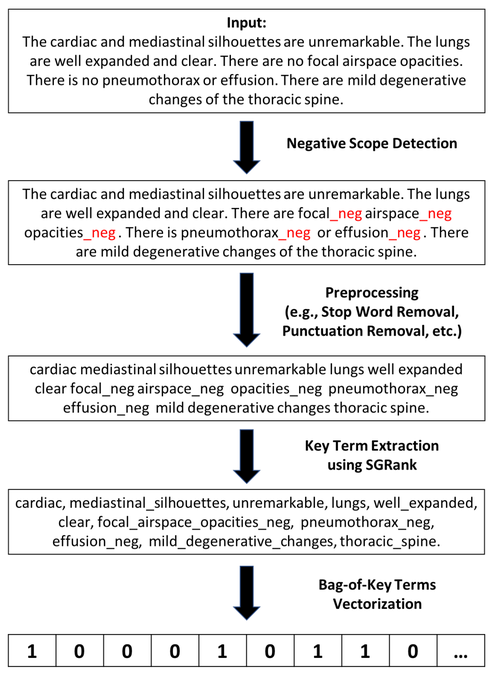
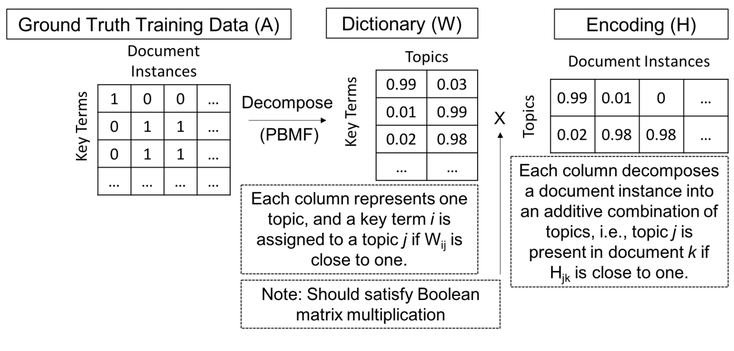


Fig: The pipeline for converting the natural language findings to a bag-of-key terms representation

Learning Topic Models using Matrix Factorization

- We propose a novel matrix factorization-based dictionary learning formulation to address the problem of learning topics from binary matrices which represent documents as bags-of-key terms.
- Assume we have a ground-truth binary matrix A where each row represents a key term and each column represents a training document.
- Element A_{ij} is 1 if key term i is present in document instance j and 0, otherwise.
- To learn an initial dictionary of topics, we decompose A into two *approximately* binary matrices:
  1. W: a dictionary of topics where each column represents one topic and each row represents a key term
  2. H: an encoding matrix that tells us how to decompose a document as an additive combination of topics



Incorporating Topic Modeling into Convolutional Neural Networks

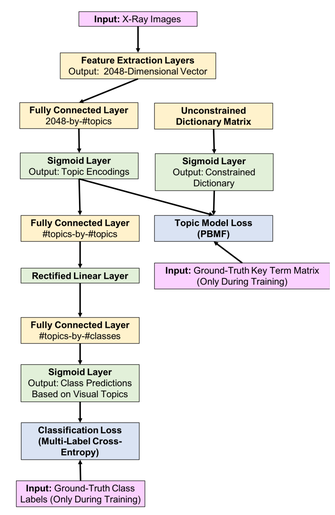


Fig: The proposed network architecture

- In practice, we need to recover the encoding matrix H from visual data, and we should adapt the topic dictionary W to account for noisy estimation of H.
- We propose a network which:
  1. Maps visual data to the topic model encoding space
  2. Finetunes the topic dictionary based on noisy visual topic recognition
  3. Uses the predicted topic encodings as interpretable features for performing x-ray classification

Quantitative Results

	Standard	Our Approach
macro-AUROC	0.857	0.867
mAP	0.459	0.477
Coverage	0.825	0.851
mIrAP	0.927	0.925

		Standard		Our Approach	
Diagnosis	# Instances	AUROC	AP	AUROC	AP
Normal	1395	0.774	0.647	0.783	0.655
Atelectasis	309	0.785	0.303	0.812	0.332
Cardiomegaly	329	0.930	0.592	0.927	0.587
Effusion	148	0.926	0.514	0.921	0.535
Emphysema	110	0.868	0.241	0.892	0.276

Table: Performance of learned visual classifiers: 1) overall (top) and 2) on individual diagnostic labels (bottom). We compare a standard ResNet-22 architecture with our modified architecture that bottlenecks through the topic modeling layer.

Qualitative Results

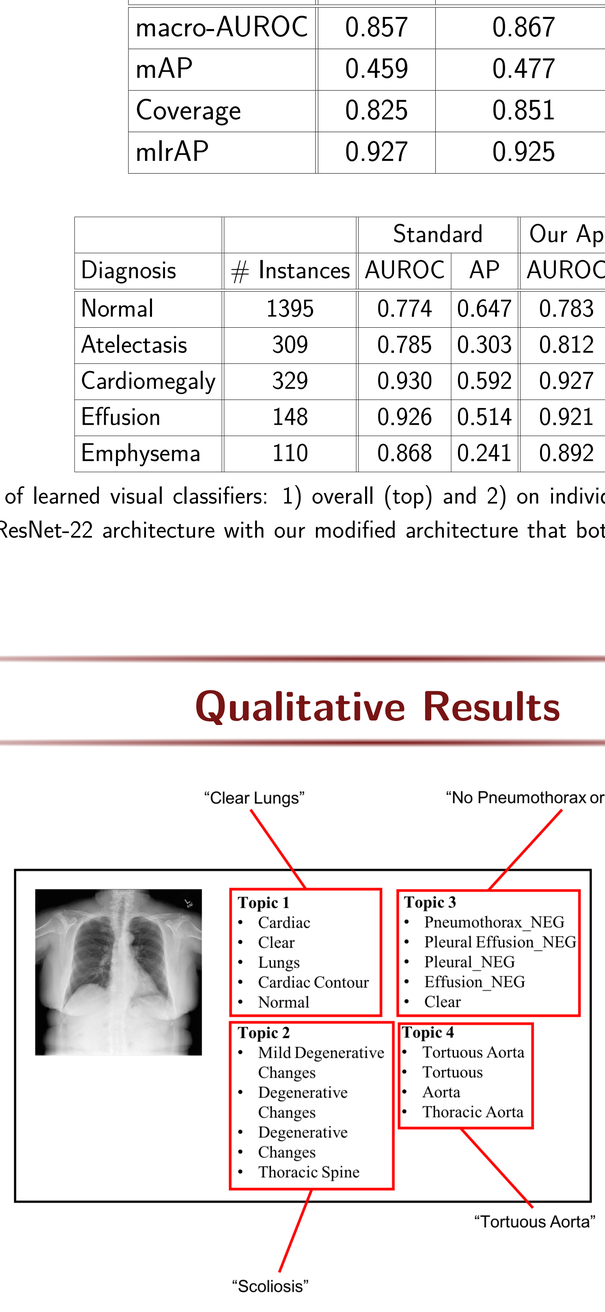


Fig: An example of an x-ray and some of its highly-ranked topics

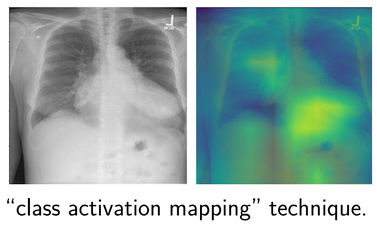


Fig: The net is also compatible with the "class activation mapping" technique. This allows it to generate attention maps for each predicted topic. This example shows where the net is attending when it predicts a topic related to "cardiomegaly".

This work is partly supported by the Air Force Office of Scientific Research (AFOSR) under the Dynamic Data-Driven Application Systems Program, NSF 1763523, 1747778, 1733843 and 1703883, and NIH 1R01HL127661-01 Awards.

IEEE International Symposium on Biomedical Imaging (ISBI), 10 April 2019, Venice, Italy