Pushing the Limits of Radiology with Joint Modeling of Visual and Textual Information

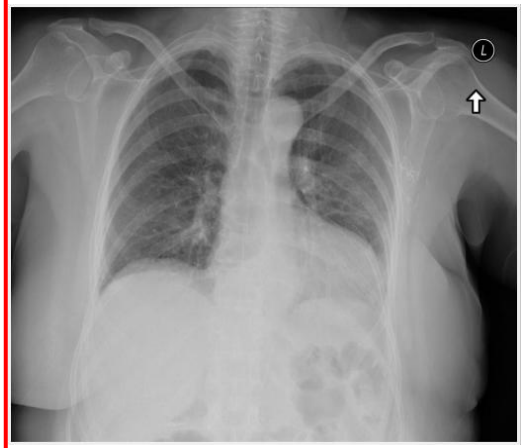
Sonit Singh
Department of Computing, Faculty of Science and Engineering, Macquarie University, Sydney NSW Australia, sonit.singh@hdr.mq.edu.au

Motivation

- Diagnostic errors play a role of up to 10% of patient deaths [1].
- 4% of radiology interpretations contain clinically significant errors [2].
- 400 million U.S. radiological studies are performed each year.
- Research studies show that the general rate of missed radiological findings can be as much as 30%.

Goal: To reduce diagnostic errors by developing clinical decision support systems that can interpret medical images and clinical text to augment radiologist's work.

A Sample Medical Image and corresponding radiology report



Comparison: PA and lateral chest redressed XXXX.
Indication: XXXX-year-old female with breast mass and smoking history.
Findings: The heart size and cardiomediastinal silhouette are normal. There is hyperexpansion of the lungs with flattening of the hemidiaphragms. There is no focal airspace opacit, pleural effusion, or pneumothorax. There multilevel degenerative changes of thoracic spine.
Impression: Emphysema, however no acute cardiopulmonary finding.
Concepts: degenerative change, emsphysema, hyperexpansion, Emphysema, Pulmonary Emphysema.

Research Questions

- Although research at the intersection of vision and language for general purpose tasks is gaining pace, its applications to healthcare are under-explored.
- Lot of medical data in the form of medical images and accompanying text reports is stored in hospital's Picture Archival and Communication Systems.
- Interpreting medical images and summarising them in natural text is challenging, complex and tedious task.

We will focus on the following research questions:

- How to automatically generate a radiology report for a given medical image?
- How can we annotate medical images from the accompanied radiology reports in a weakly supervised manner?
- How can we highlight relevant area in a medical image based on the features extracted from radiology reports?
- How can we develop cross-modal models for medical indexing and retrieval?

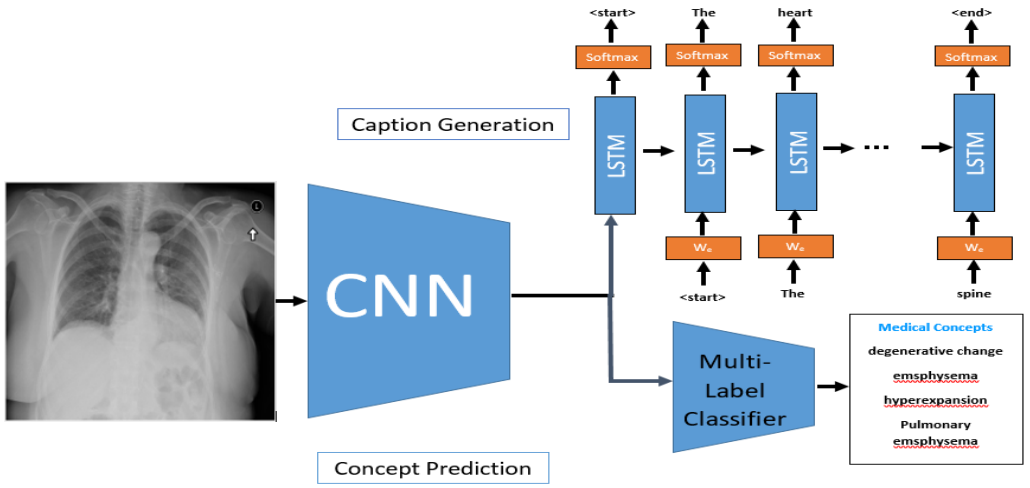
Experimental Setup

Datasets

- Various publicly available datasets: ChestX-ray8, Open-i, and ImageCLEFcaption, PEIR Radiology and pathology images, and Indiana University Chest X-ray dataset.
- Medical images and radiology reports from Macquarie University Hospital.

Evaluation Metrics

- Medical image captioning (Med-Captioning): Standard image captioning metrics, namely, BLEU, ROUGE, METEOR, CIDEr, SPICE and WMD.
- Medical Visual Question Answering (VQA-Med): Standard Accuracy metric and Wu-Palmer Similarity score.
- Medical Visual Dialog (VisDial-Med): Standard retrieval metrics, namely, recall@k and mean reciprocal rank (MRR)
- Concept Prediction: Multi-label classification evaluation metrics such as precision, recall, F-score, precision@k, and recall@k.



Experiments

Tasks

- Two tasks: (1). Caption generation, and (2). Concept prediction (MeSH terms in the medical images)

Dataset

- Indiana University Chest X-ray dataset.
- Consists of 7,470 medical images and 3,955 radiology reports.
- Major sections in : parentImage_id, comparison, indication, findings, impression and MeSH terms.

Approach

Caption Prediction

- The main approaches are: (1). Using *templates* that rely on detectors and map the output to linguistic structures. (2). Using *language models* based on Recurrent Neural Networks, and (3). Using *retrieval* based methods.
- We followed encoder-decoder based approach.
- Used VGG16 for image features detection pre-trained on ImageNet, and LSTM for language generation.

Concept Prediction

- Each report has key concepts (MeSH terms) which are present in a Chest X-ray.
- Concepts prediction is a *multi-label classification*.
- Visual features are extracted using VGG16, and then these features are passed through multi-label classifier.

Preliminary Results

Table I: Caption generation results on IU ChestX-ray dataset in terms of BLEU (B-n), CIDEr (C), METEOR (M), and ROUGE (R) scores.

B-1	B-2	B-3	B-4	C	M	R
0.19	0.05	0.02	0.01	0.30	0.07	0.33

Table II: Concept prediction results on IU ChestX-ray dataset in terms of recall@k values

recall@5	recall@10	recall@20
0.2514	0.3375	0.4852

Conclusion and Future Work

- We argue the need for integrating language and vision processing in the medical domain, and provide various research directions.
- To improve results, we will use CNN-RNN model for Concept prediction, combining deeper CNN models and Stacked LSTM for Caption generation.

Acknowledgements
I am thankful to my supervisors Sarvnaz Karimi, Kevin Ho-Shon (Director of Radiology, Macquarie University Hospital) and Len Hamey for their valuable suggestions. I am also thankful to Google for their generous support by providing travel grant to attend the conference.

References

[1]. National Academy of Medicine. Improving diagnosis in Health Care. The National Academic Press; 2015.
[2]. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D, "Interpretive Error in Radiology", American Journal of Roentology; 2016; 1-11.
[3]. B. Jing, P. Xie, and E. Xing, "On the Automatic Generation of Medical Imaging Reports", Association for Computational Linguistics, July, 2018, Melbourne, Australia.