# **On the Automatic Generation of Medical Imaging Reports**

### Introduction

### Motivation

Medical imaging is widely used in clinical practice for diagnosis and treatment. Report-writing can be error-prone for unexperienced physicians, and time-consuming and tedious for experienced physicians. To address these issues, we study the automatic generation of medical imaging reports.



Impression: No acute cardiopulmonary abnormality.

Findings: There are no focal areas of consolidation. No suspicious pulmonary opacities. Heart size within normal limits. No pleural effusions. There is no evidence of pneumothorax. Degenerative changes of the thoracic spine.

MTI Tags: degenerative change

Figure 1: Example of Medical Report for a Chest X-ray Image

# Challenges

(1) A report contains multiple heterogeneous forms of information, including *findings* and *tags*. (2) Abnormal regions in medical images are difficult to identify.

(3) The reports are typically long and contain many sentences.

### Contributions

(1) We build a multi-task learning framework which jointly performs the prediction of tags and the generation of paragraphs.

(2) We propose a co-attention mechanism to localize regions containing abnormalities and generate narrations for them.

(3) We develop a hierarchical LSTM model to generate long paragraphs.

(4) We perform extensive experiments to show the effectiveness of the proposed method.

### **Proposed Model**



Figure 2: Overview of the Proposed Model

# **Encoding Process**

**Visual Information**: (1) We use a CNN to learn visual features for different sub-regions of a given image. (2) These visual features are fed into a *multi-label classification* (MLC) network to predict relevant tags. **Semantic Information**: (1) Each tag is represented by a word-embedding vector. (2) The wordembedding vectors of tags serve as the semantic features of this image. Mix Visual & Semantic Information The visual and semantic features are fed into a *co-attention* model to generate a context vector that simultaneously captures the visual and semantic information.

# Decoding Process

**Sentence LSTM**: The context vector is input into the sentence LSTM, which produces topic vectors through topic generator and controls the termination through stop control. (2) Word LSTM: Given a topic vector, the word LSTM takes it as input and generates a sequence of words to form a sentence.

### Datasets

Dataset Description			Dataset	Methods	B-1	B-2	B-3	B-4		Rouge	
X-Ray A	set of chest x-ray images	paired with their		CNN-RNN[1]	0.316	0.211	0.140	0.095	0.159	0.267	0.111
corresponding diagnostic reports. The dataset contains 7,470 pairs of images and reports.			X-Ray	LRCN[2]	0.369	0.229	0.149	0.099	0.155	0.278	0.190
				Soft ATT[3] ATT-RK[4]	$0.399 \\ 0.369$	$0.251 \\ 0.226$	$0.168 \\ 0.151$	$0.118 \\ 0.108$	$0.167 \\ 0.171$	$0.323 \\ 0.323$	$0.302 \\ 0.155$
PEIR The PEIR Gross dataset contains 7,442 image				Ours-no-Attention	0.505	0.220 0.383	0.131 0.290	0.108 0.224	0.171 0.200	$\frac{0.323}{0.420}$	$0.155 \\ 0.259$
caption pairs from 21 different sub-categories.				Ours-Semantic-only		0.371	0.290	0.221	0.200	0.418	0.286
Table 1: Dataset descriptions.				Ours-Visual-only	0.507	0.373	0.297	0.238	0.211	0.426	0.300
				Ours-CoAttention	0.517	0.386	0.306	0.247	0.217	0.447	0.327
Tag Prediction				CNN-RNN[1]	0.247	0.178	0.134	0.092	0.129	0.247	0.205
				LRCN[2]	0.261	0.184	0.136	0.088	0.135	0.254	0.203
				Soft ATT[3]	0.283	0.212	0.163	0.113	0.147	0.271	0.276
				ATT-RK[4]	0.274	0.201	0.154	0.104	0.141	0.264	0.279
Dataset	Methods R@5	R@10 R@20		Ours-No-Attention					0.131		
		$\begin{array}{c c c c c c c c c c c c c c c c c c c $		Ours-Semantic-only							0.274
A-May				Ours-Visual-only					0.149		0.280
	Ours-CoAttention 0.644			Ours-CoAttention							
PEIR	VGG-19 0.392			Table 3: Main r				· · · ·			
	Ours-CoAttention <b>0.398</b>	0.494 <b>0.596</b>		dataset (upper part), and single sentence generation on PEIR Gross dataset (lower part). BLUE-n denotes the BLEU score							
ble 2: Tag p	prediction on IU X-Ray and	Gross dataset (Ic		art). E	SLUE-r	n deno	tes the	BLEU	score		

R denotes recall.

Baoyu Jing<sup>1,2</sup> Pengtao Xie<sup>1</sup> Eric P. Xing<sup>1</sup>

<sup>1</sup>Petuum Inc. <sup>2</sup>Carnegie Mellon University

# Main Results

uses up to n-grams.





mid-thoracic vertebral body.

- attention. ICML

### Paragraph Generation

changes of the spine.

ventration of the right hemidiaphr

pleural effusion or pneumothorax.

No focal airspace consolidation. No

hernia. No focal consolidation pleural effusion or pneumothorax. No acute bony

abnormality

and expanded. Heart and mediasting

### Ours-CoAttention

ctive disease. The heart and in the interval. Lungs are clear led. Cardiomediastinal silhouette within normal limits. No pleural ffusion or pneumothorax is seen. No leural effusion. No cavitary or neumothorax

limits. There are multilevel degenerative onsolidation. Degenerative changes he spine. This is moderate exam of the pneumothorax. Lungs are clear There is no focal airspace consolidation pleural effusion or pneumothorax.

> No acute cardiopulmonary abnormal lower right midlung. No leural effusion or focal airspace disea osclerotic calcifications bor ructures bilaterally. There is no pleura sion or pneumothorax developed he right lower lobe.

> ions. The lungs are otherwise rossly clear. Resolution of by normal leural effusion

### Figure 3: Examples of generated paragraphs. **Co-Attention Learning**



Ground truth paragraph: No cute cardiopulmonary fin cardiopulmonary silhouette is normal. There is no focal airspace opacity pleural effusior or pneumothorax. The obstruction are intact with mild degenerative change in the

Generated paragraph: (1) No acute intrathoracic abnormality (2) No bony abnormality. The cardio mediastinal silhouette is within normal limits for appearance. (4) No focal areas of pulmonary consolidation. (5) Breast motion. (6) There is an age indeterminate deformity of a



scarring; sternotomy

**Generated Sentence** (4):

normal; degenerative change; nodule; 🛛 yper expansion ranulomatous disease anuloma; pneumonia

arring; sternotomy



scarring; sternotomy



normal; <u>degenerative change;</u> nodule; c hyper expansion; granulomatous disease granuloma; pneumonia scarring; sternotomy



normal; nodule;

scarring; sternotomy Generated Sentence (6)



granulomatous disease granuloma; pneumonia scarring; sternotomy

### Figure 4: Co-Attention Learning.

### References

[1] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. CVPR

[2] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. CVPR

[3] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual

[4] Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. 2016. Image captioning with semantic attention. CVPR