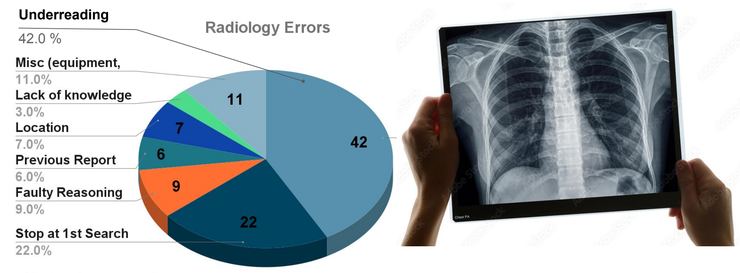
RepsNet: Combining Vision with Language for Automated Medical Reports

Ajay Tanwani, Joelle Barral, Daniel Freedman - Verily and Google Research

Introduction

- 1 billion radiology examinations are performed worldwide with 3-5 % error rate

- Practitioners spend 5-10 minutes for writing each report

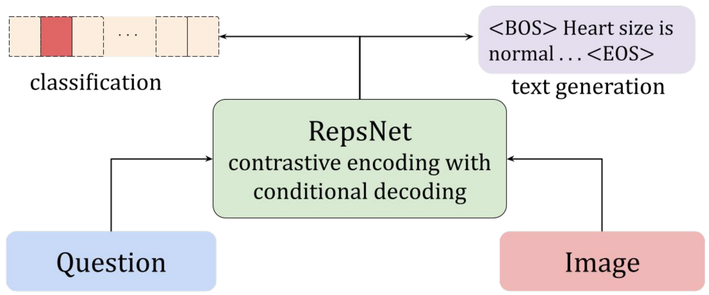


Objective

to improve report accuracy and save time in writing and explaining reports to patients, while automating billing codes for describing medical findings

RepsNet Main Idea

- RepsNet fills medical reports by **classification** or **text generation** via visual question answering



Q: What type of image modality is used ?
A: X-ray

Q: Which plane is the image taken in ?
A: PA (Posterior-Anterior)

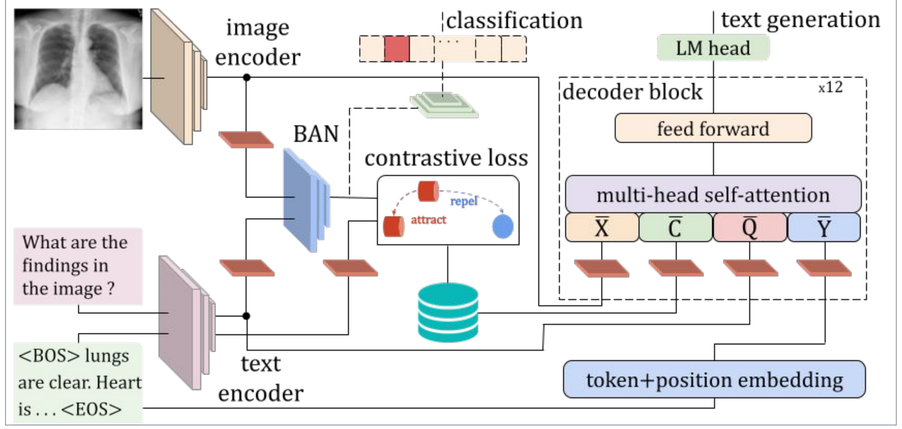
Q: What are the findings ?
A: Aorta is tortuous and calcified. There are degenerative changes of acromioclavicular joints and spine.

RepsNet Model

- RepsNet adapts pre-trained vision and language models

- Joint pre-training of images and text in the encoding phase by contrastive learning

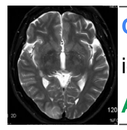
- Augment language model with image and prior context of nearest neighbour answers in the decoding phase



VQA-Rad Results

- State-of-the-art classification accuracy by adapting RepsNet on small VQA-Rad datasets

Question: Where is the cystic lesion in the CT?
Answer: Left kidney



Question: Are the ventricles visualized in MRI?
Answer: No

	2018	2019	All
MEVF [5]	66.10	–	–
MMQ [11]	67.00	–	–
QCR [41]	69.65	–	–
CLEF [1]	–	62.40	–
CRPD [21]	72.70	–	–
RepsNet-0	**81.08**	**67.57**	**63.69**
RepsNet-5	**83.55**	**79.83**	**71.93**
RepsNet-10	**87.05**	**81.17**	**80.37**

IU-XRay Results

- Joint pre-training of images and medical findings in the encoding phase, and conditional decoding of GPT-2 language model on image and nearest neighbour reports gives state-of-the-art BLUE scores on IU-Xray dataset

	B1	B2	B3	B4
Co-Att [16]	0.45	0.29	0.20	0.15
HRGR [20]	0.44	0.30	0.21	0.15
CMAS [15]	0.46	0.30	0.21	0.15
Mem-T [7]	0.47	0.30	0.22	0.16
VTI [25]	0.49	0.36	0.29	0.15
PPKED [22]	0.48	0.31	0.22	0.17
RepsNet	**0.58**	**0.44**	**0.32**	**0.27**

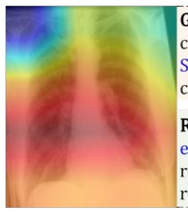
Qualitative Results & Next Steps

- RepsNet shows strong alignment with ground-truth in describing medical findings

- Ongoing work: Grounding medical knowledge + video report generation + application to gastroenterology

GroundTruth: Cardiomediastinal silhouette is normal in size and configuration. Pulmonary vasculature within normal limits. Lungs are well aerated. There is no pneumothorax, pleural effusion or focal consolidation.

RepsNet: Heart size is normal. Lungs are clear. There are no focal airspace consolidations. No pleural effusions or pneumothoraces. Hilar and mediastinal contours are normal. Normal pulmonary vascularity.



GroundTruth: Sternotomy and surgical clips are again seen. Cardiomediastinal contours are unchanged. There is background of marked centrilobular emphysema. Streaky opacities in lung bases may represent atelectasis or scarring. There is no consolidation, pleural effusion or pneumothorax.

RepsNet: Hyperexpanded lungs with flattened hemidiaphragms consistent with emphysema. There is streaky airspace opacities in left suprahilar and lingular regions. No pneumothorax or effusions. Mild bilateral costophrenic blunting represents pleural thickening and scarring. Degenerative changes of thoracic spine.

contact: <ajaytanwani@google.com>

website: <https://sites.google.com/view/repsnet>