Complex Organ Mask Guided Radiology Report Generation (Supplementary Material)

Tiancheng Gu, Dongnan Liu, Zhiyuan Li, Weidong Cai
University of Sydney, Sydney, Australia
{tigu8498, zhli0736}@uni.sydney.edu.au, {dongnan.liu, tom.cai}@sydney.edu.au

1. Methodology

1.1. Details on Tissue Segmentation Masks

We provide more details regarding the tissue segmentation masks extracted by the pre-train CXAS model [5] for the proposed Mask-guided Organ Prototype Feature Extraction mechanism (Sec. 3.1). The details are presented in Table 1. For each given image, the CXAS model can extract totally a fixed number of masks for different tissues in the radiology images, such as lung lobes, lung zones, and lung halves. However, some tissues are irrelevant to the descriptions in the report, such as the abdomen, trachea, etc. Therefore, out of a total of 159 generated mask images, only 97 are useful for the COMG model. These include 70 for bones, 15 for lungs, 6 for the heart, and 6 for the mediastinum. Fig. 1 shows the example of mask images extracted from the input image using the CXAS model.

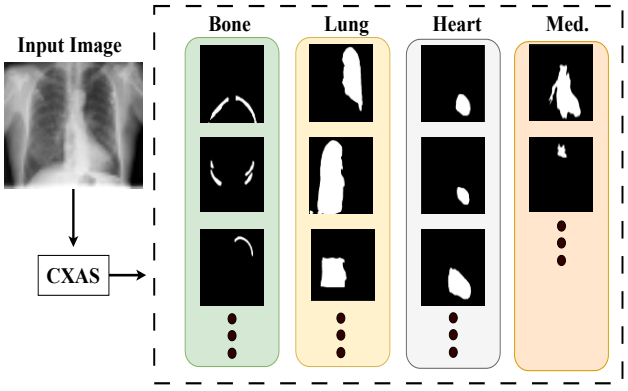

Figure 1. The mask image generated after the CXAS from the MIMIC-CXR benchmark. Med. means the mediastinum.

1.2. Disease Symptom Graph

We have presented the details on the disease knowledge graph obtaining the prior-disease captions, as presented in Fig. 2. It is based on the professional analysis of the relationship between the organs and the corresponding diseases in the radiology images. Specifically, the graph was developed in [3], taking into account symptom correlation, symptom characteristics, occurrence location, and other relevant factors.

Regional Mask Images	Num.	Region	Total Mask
Lung lobes	5	Lung	159
Lung zones	8		
Lung halves	2		
Heart region	6	Heart	
Mediastinum	6	Mediastinum	
Ribs	46	Bone	
Ribs super	24		
...	...	...	

Table 1. The specific information of masks generated by the CXAS model [5], as well as the mask images we ultimately used.

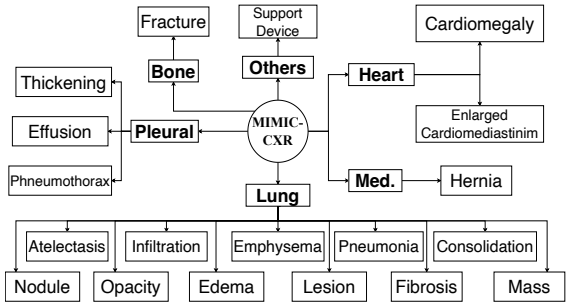

Figure 2. The symptom graph summarizes the related diseases for each organ in the MIMIC-CXR dataset. Specifically, we focused on four organs: bone, lung, heart, and mediastinum.

2. Dataset

Table 2 provides detailed information on the two benchmarks used to evaluate the COMG model. The table shows that MIMIC-CXR contains more cases than IU-Xray. Addi-

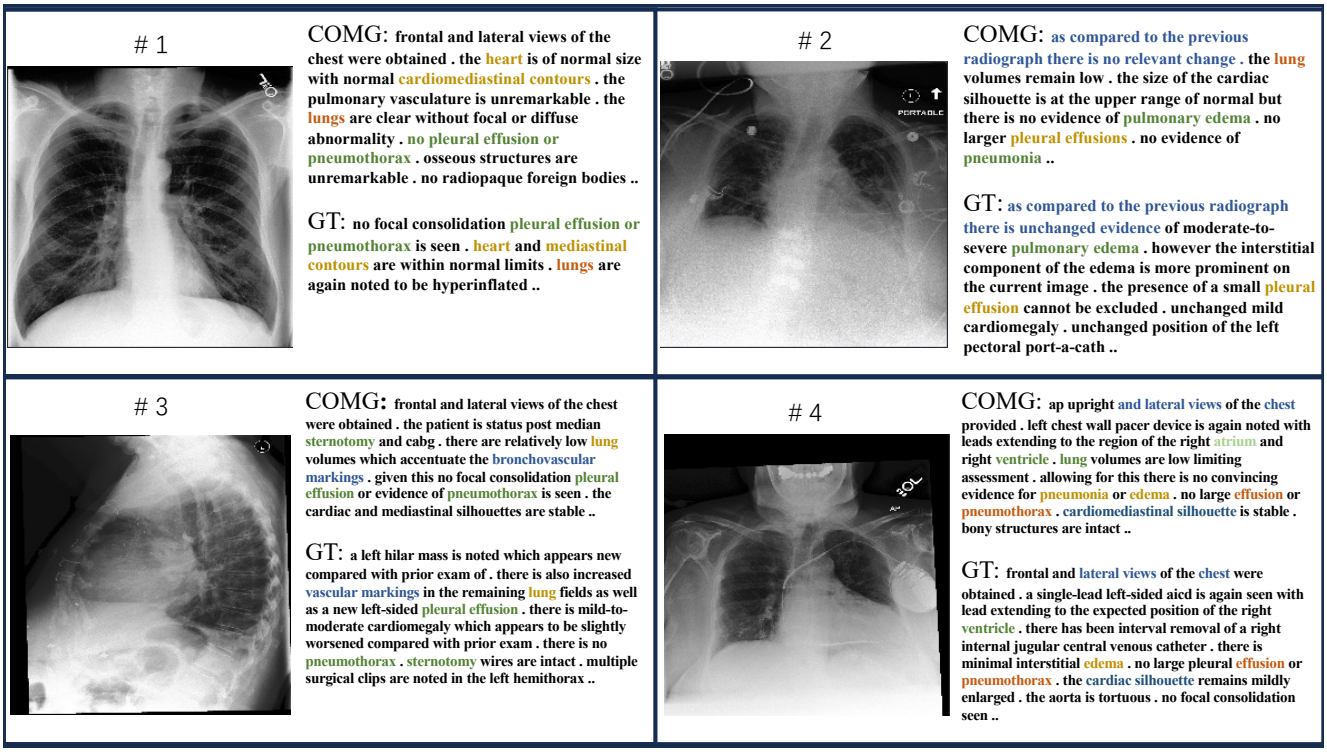

Figure 3. The visulization of prediction results by the COMG model. GT is the abbreviation of the Ground Truth.

Dataset	IU-Xray [2]			MIMIC-CXR [4]		
	Train	Val.	Test	Train	Val.	Test
Image	5.2K	0.7K	1.5K	369.0K	3.0K	5.2K
Report	2.8K	0.4K	0.8K	222.8K	1.8K	3.3K
Patient	2.8K	0.4K	0.8K	64.6K	0.5K	0.3K
Avg. Len.	37.6	36.8	33.6	53.0	53.1	66.4

Table 2. The specifications of two benchmark datasets that will be utilized to test the COMG model.

tionally, the table provides detailed information on how the data is split for these two benchmarks [1].

3. Visualization Results

This section demonstrates more visualization results predicted by the COMG model. More details are shown in Fig. 3. We have found that the reports generated by our model are more closely related to the ground truth. This includes key organs such as the heart, lungs, and atrium, as well as corresponding diseases such as pleural effusions, pneumothorax, etc.

References

[1] Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan. Generating radiology reports via memory-driven transformer. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, Aug. 2021. 2

[2] Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2016. 2

[3] Zhongzhen Huang, Xiaofan Zhang, and Shaoting Zhang. Kiut: Knowledge-injected u-transformer for radiology report generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19809–19818, 2023. 1

[4] Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317, 2019. 2

[5] Constantin Seibold, Alexander Jaus, Matthias A Fink, Moon Kim, Simon Reiß, Ken Herrmann, Jens Kleesiek, and Rainer Stiefelhagen. Accurate fine-grained segmentation of human anatomy in radiographs via volumetric pseudo-labeling. *arXiv preprint arXiv:2306.03934*, 2023. 1