Exploring and Distilling Posterior and Prior Knowledge for Radiology Report Generation

Fenglin Liu[1], Xian Wu[2], Shen Ge[2], Wei Fan[2], Yuexian Zou[1,3]

[1] ADSPLAB, School of ECE, Peking University

[2] Tencent Medical AI Lab, Beijing, China   [3] Peng Cheng Laboratory, Shenzhen, China

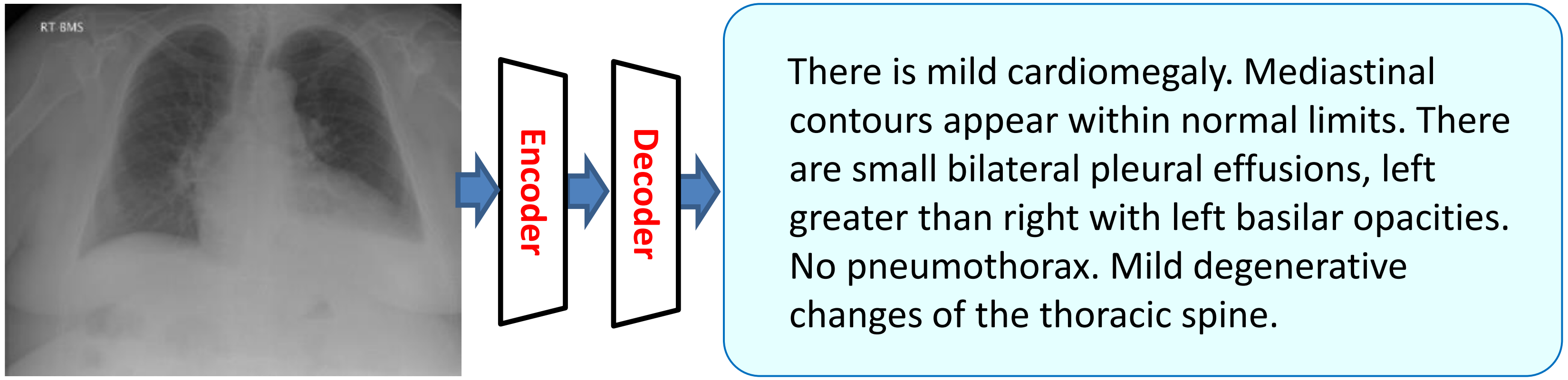
北京大學 PEKING UNIVERSITY

Contents

1. Introduction

Radiology Report Generation

- Dataset: (V,S), where V and $S=\{s_1, s_2, \ldots, s_T\}$ represent the input radiology image and the target report, respectively.
- Encoder-Decoder Framework: In the encoding stage, the global image features are extracted by CNN from the entire image; In the decoding stage, the whole report is generated using HRNN.
- Training Objective: The widely-used training objective is to minimize the cross entropy loss.

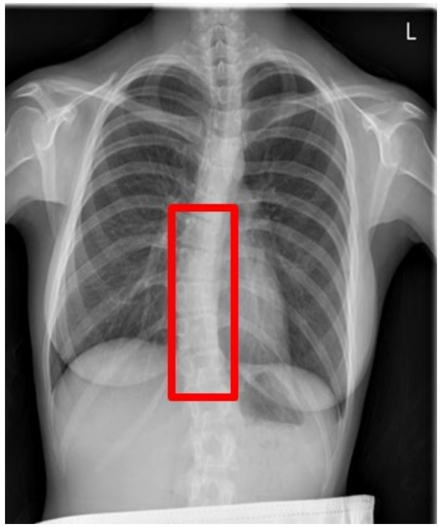


$$\text{Visual Enc.}: \mathcal{V} \rightarrow \hat{\mathcal{V}}; \quad \text{Target Dec.}: \hat{\mathcal{V}} \rightarrow \mathcal{S}.$$

$$L_{CE}(\theta) = -\sum_{t=1}^{T} \log\left(p_\theta\left(s_t^* | s_{1:t-1}^*; \mathcal{V}\right)\right)$$

PEKING UNIVERSITY

Motivations: Visual Data Deviation



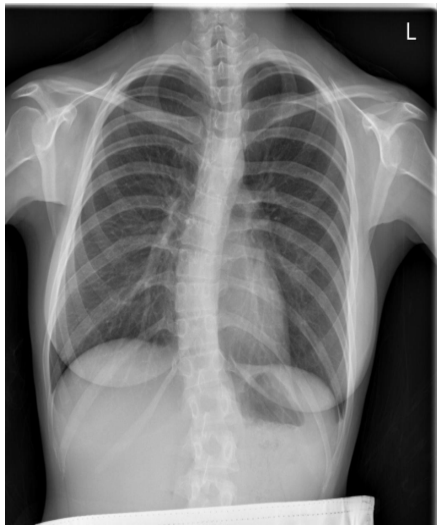
Medical Report:

Lungs are clear. No pleural effusions or pneumothoraces. Heart and mediastinum of normal size and contour. scoliosis.

- The normal images dominate the dataset over the abnormal ones [1], especially for the rare diseases.
- For each abnormal image, the appearance of abnormal regions (red bounding box) only occupy a small part of the entire image.
- As a result, this unbalanced visual distribution would distract the model from accurately capturing the features of rare and diverse abnormal regions

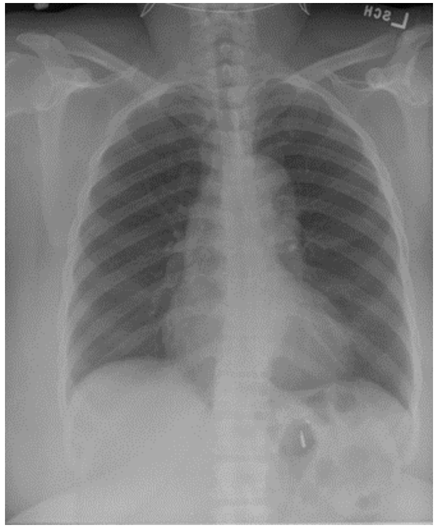
[1] Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. In CVPR, 2016.

Motivations: Textual Data Deviation



Medical Report:

Lungs are clear. No pleural effusions or pneumothoraces. Heart and mediastinum of normal size and contour. scoliosis.



Medical Report:

The heart size is enlarged. The aorta is tortuous. The pulmonary vasculature appears normal. Lungs are otherwise clear bilaterally. No pleural effusions or pneumothorax. No bony abnormalities.

- In a report, radiologists tend to describe all the items in an image, making the description of abnormal regions (red colored text) only occupy a small part of the entire report.
- Besides, there are many similar sentences (blue colored text) used in each report to describe the normal regions.
- With this unbalanced textual distribution, training with such dataset makes the generation of normal sentences dominant, disabling the model to describe specific crucial abnormalities.

北京大學 PEKING UNIVERSITY

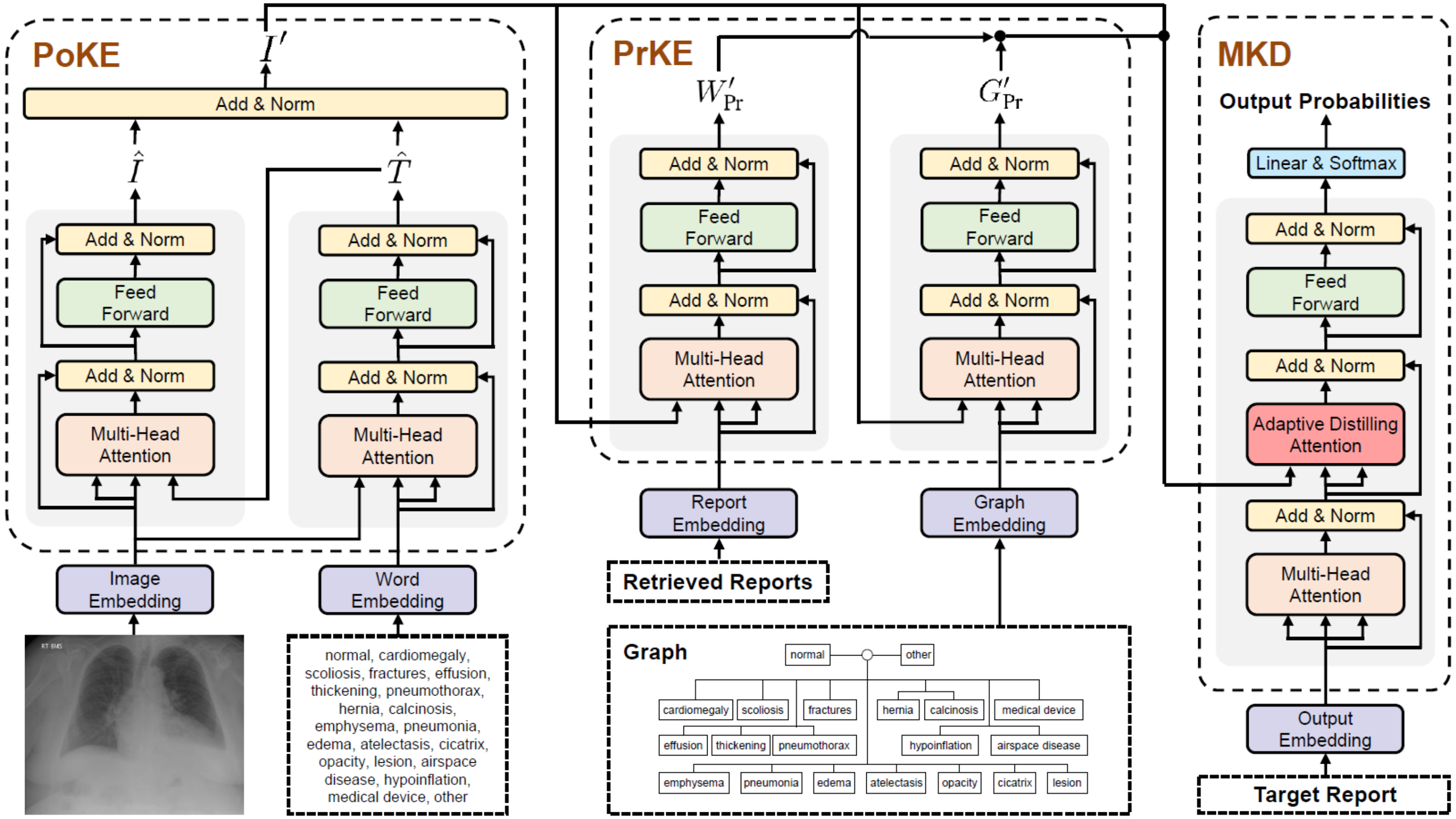
2. Approach: Posterior-and-Prior Knowledge Exploring-and-Distilling

Overview: PPKED

- The Posterior-and-Prior Knowledge Exploring-and-Distilling (PPKED) imitates the radiologists' working patterns. Given a medical image, radiologists will:
  - 1. examine the abnormal regions and assign the disease topic tags to the abnormal regions;
  - 2. then accurately write a corresponding report based on years of prior medical knowledge and prior working experience accumulations.

Overview: PPKED

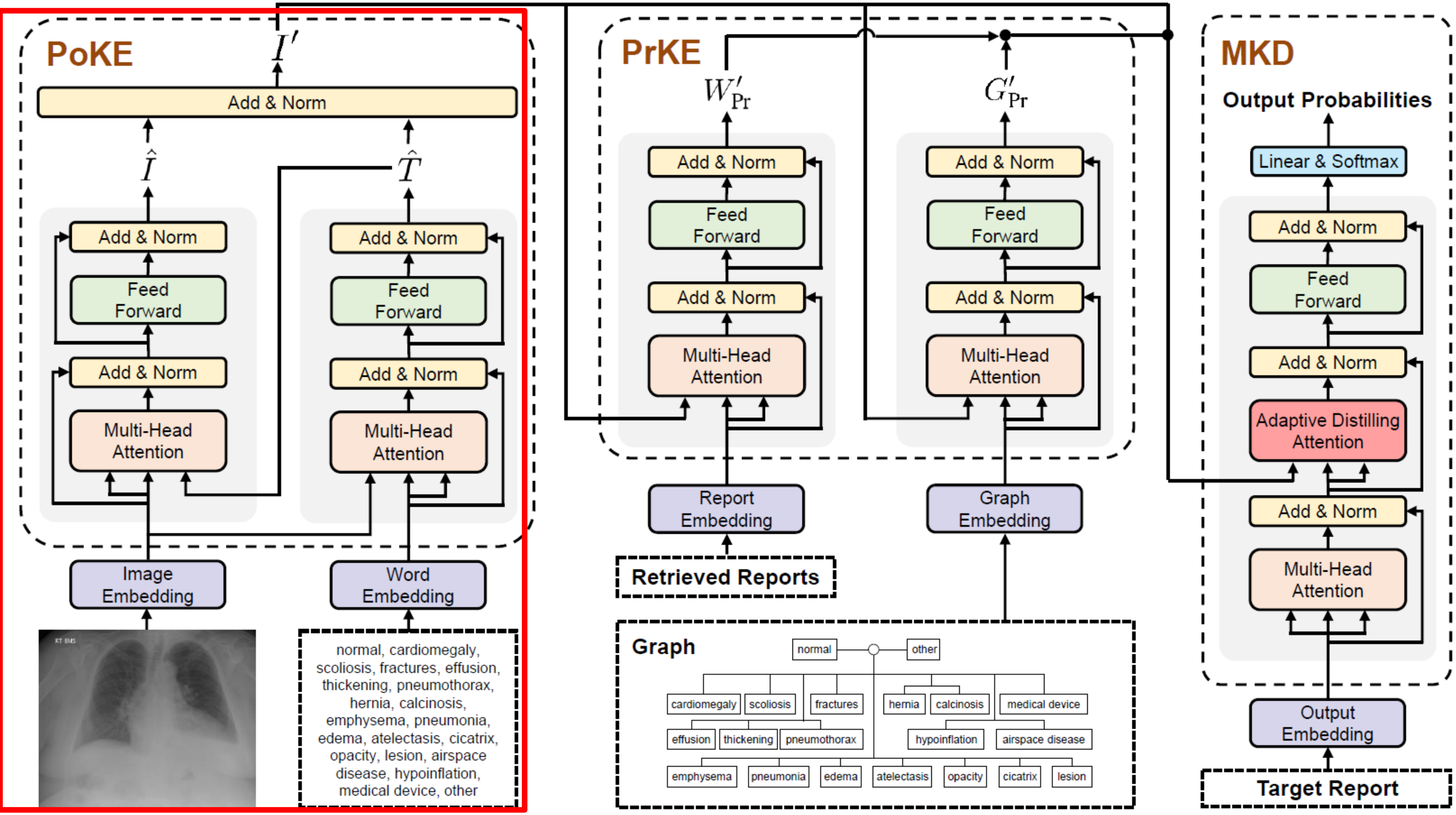
- To model above working patterns, the PPKED introduces Posterior Knowledge Explorer (PoKE), Prior Knowledge Explorer (PrKE) and Multi-domain Knowledge Distiller (MKD).



Our approach based on the Multi-Head Attention (MHA) and Feed-Forward Network (FFN) from Transformer [1].

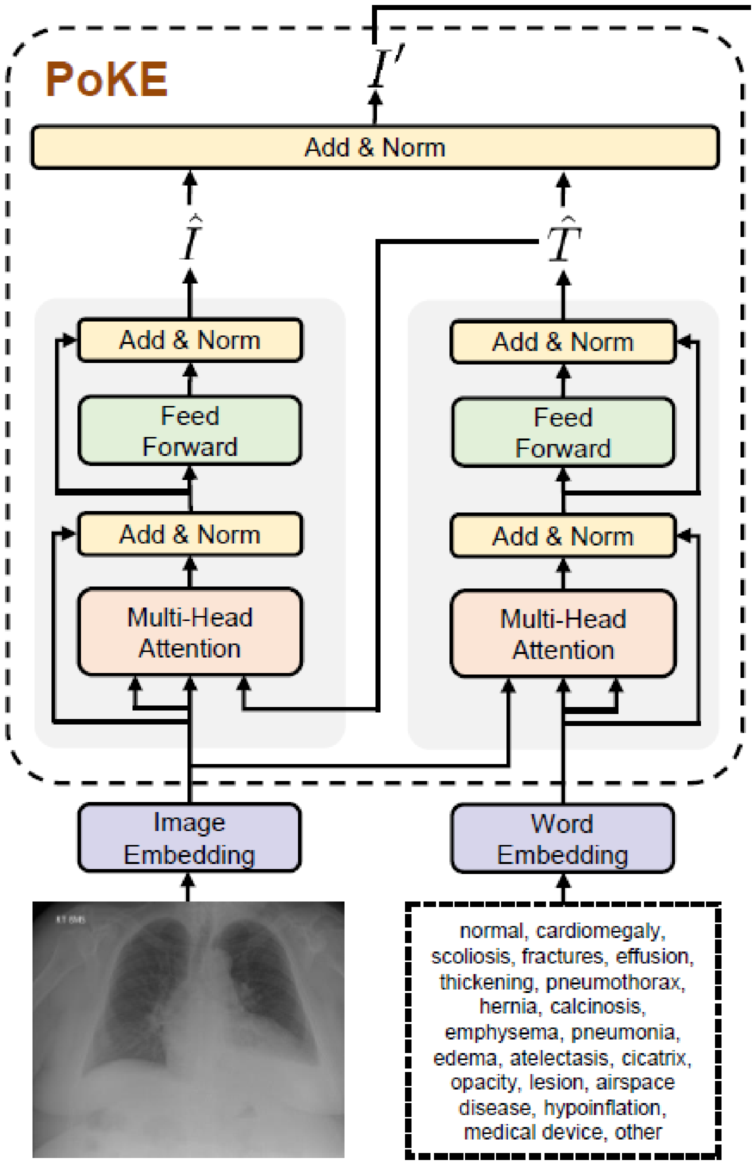
[1] Attention is all you need. In NIPS, 2017.

Posterior Knowledge Explorer (PoKE)



It is designed to alleviate visual data deviation by extracting the abnormal regions based on the input image.

Posterior Knowledge Explorer (PoKE)



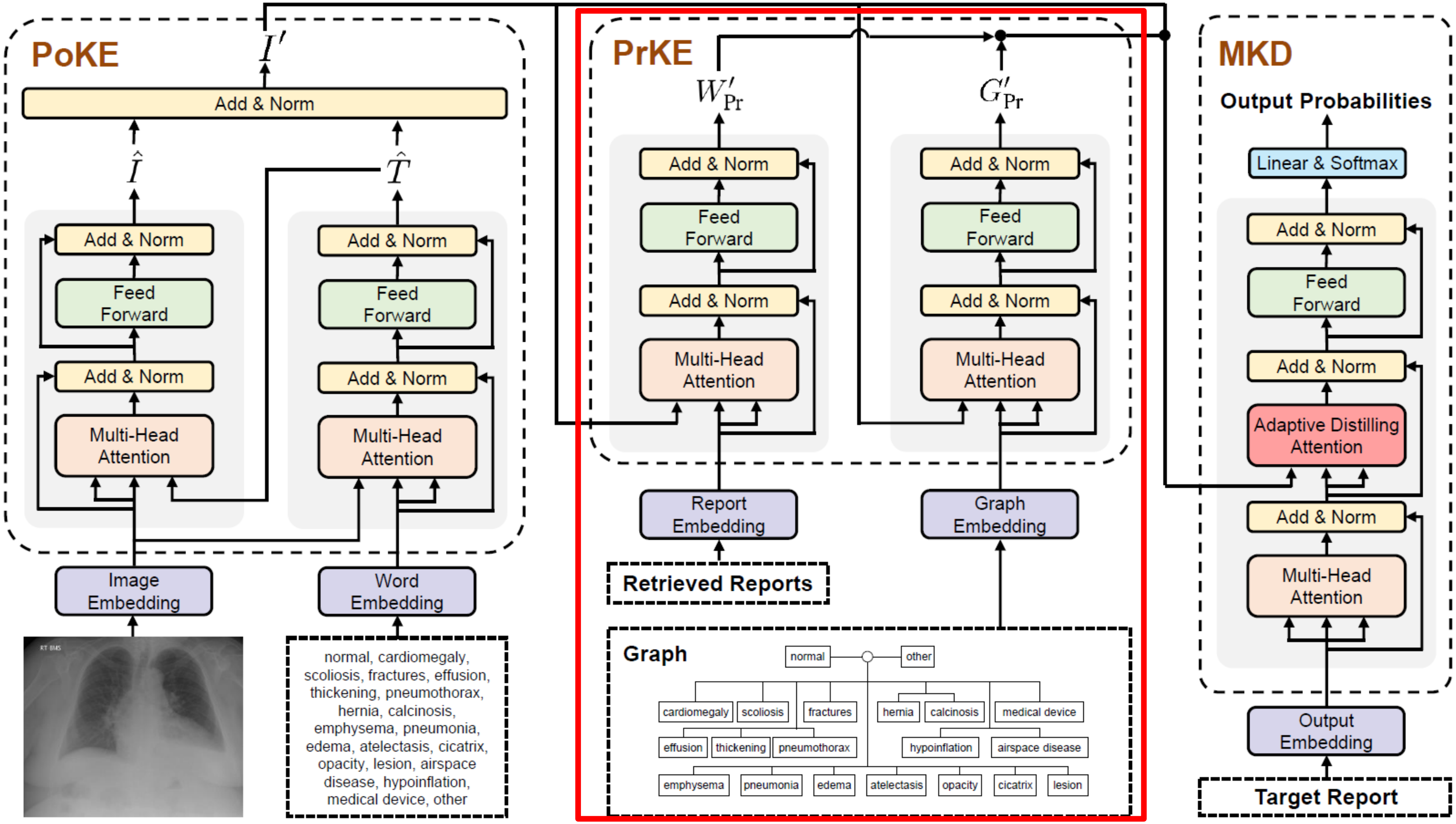
Given the input image I and disease topics tags T:

$$\hat{T} = \mathrm{FFN}(\mathrm{MHA}(I, T)); \ \hat{I} = \mathrm{FFN}(\mathrm{MHA}(\hat{T}, I))$$

$$I' = \mathrm{LayerNorm}(\hat{I} + \hat{T})$$

i.e., the I are first used to find the most relevant topics and filter out the irrelevant topics. Then the attended topics $\hat{T}$ are further used to mine topic related image features $\hat{I}$.

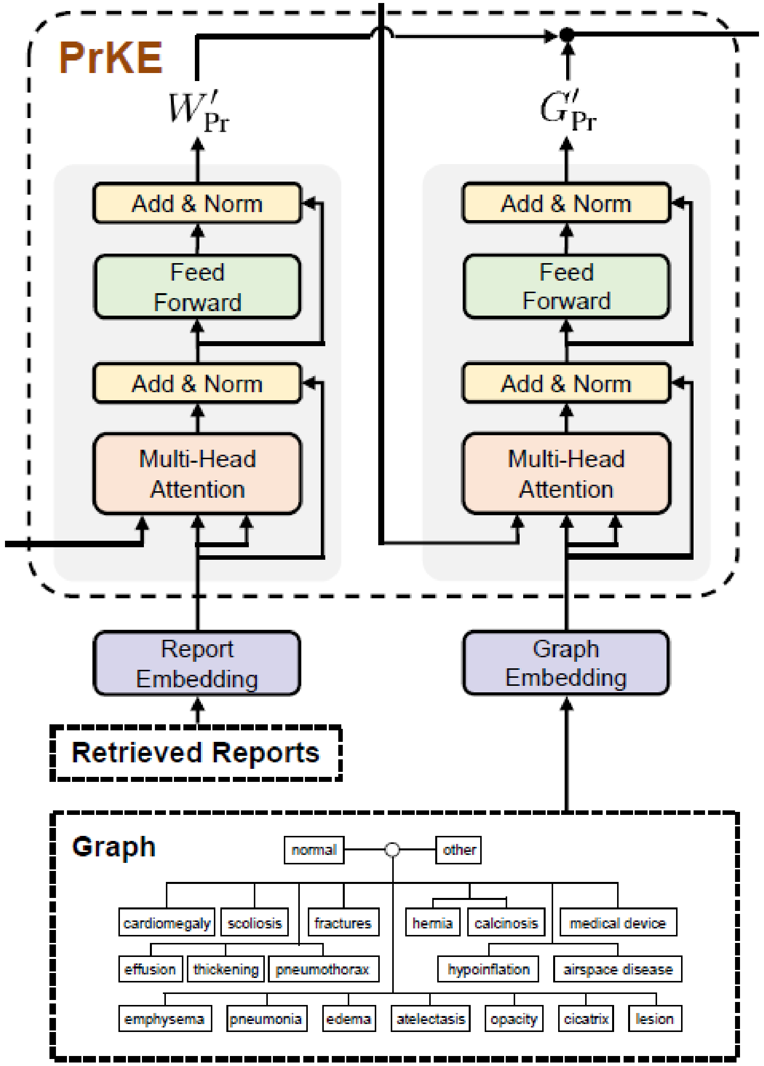
Prior Knowledge Explorer (PrKE)



The PrKE is designed alleviate textual data deviation by encoding the prior knowledge.

北京大学
PEKING UNIVERSITY

Prior Knowledge Explorer (PrKE)



The prior knowledge includes the prior radiology reports W_{Pr} (i.e., prior working experience) pre-retrieved from the training corpus and the prior medical knowledge graph G_{Pr} (i.e., prior medical knowledge), which models the domain-specific prior knowledge structure and is pre-constructed from the training corpus:

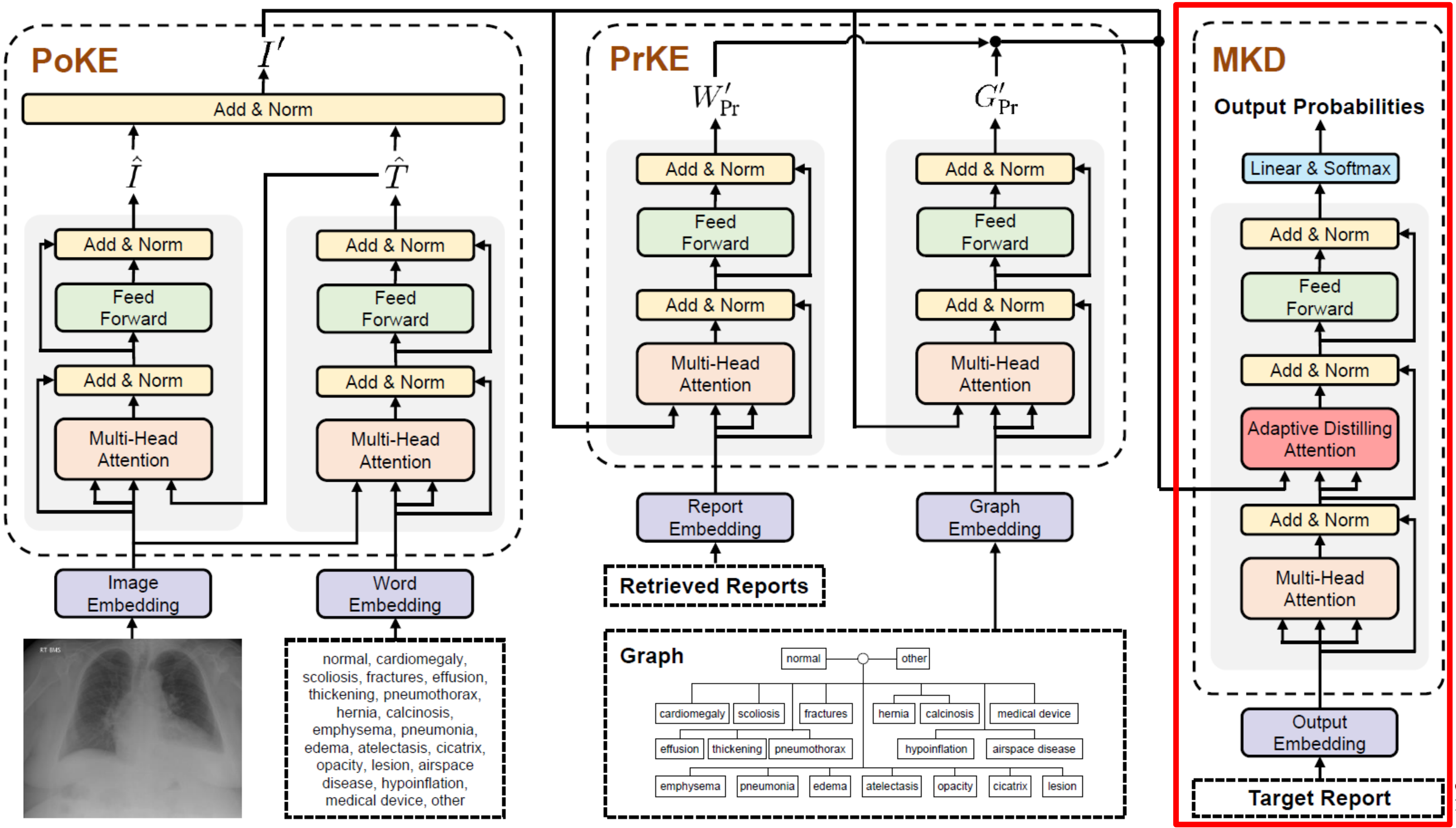
$$W'_{\mathrm{Pr}} = \mathrm{FFN}(\mathrm{MHA}(I', W_{\mathrm{Pr}}))$$

$$G'_{\mathrm{Pr}} = \mathrm{FFN}(\mathrm{MHA}(I', G_{\mathrm{Pr}}))$$

By processing I' through these two equations, we can acquire W'_{Pr} and G'_{Pr} which represent the prior knowledge relating to the abnormal regions I' of the input image.

PEKING UNIVERSITY 北京大學

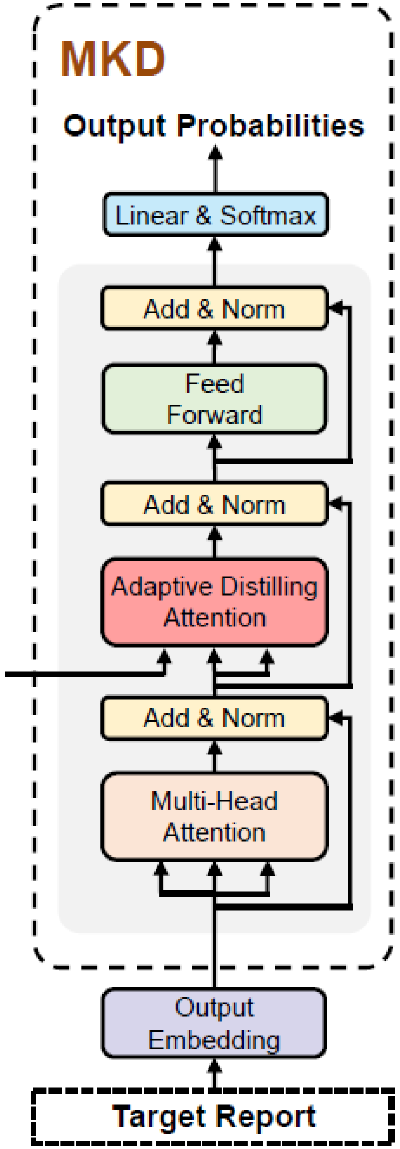
Knowledge Distiller (MKD)



Finally, the MKD is designed to distill the useful knowledge to generate proper reports.

北京大学 PEKING UNIVERSITY

Knowledge Distiller (MKD)



Given the embedding of current input word x_t:

$$h_t = \mathrm{MHA}(x_t, x_{1:t})$$

$$h'_t = \mathrm{ADA}(h_t, I', G'_{\mathrm{Pr}}, W'_{\mathrm{Pr}})$$

$$\mathrm{ADA}(h_t, I', G'_{\mathrm{Pr}}, W'_{\mathrm{Pr}}) = \mathrm{MHA}(h_t, I' + \lambda_1 G'_{\mathrm{Pr}} + \lambda_2 W'_{\mathrm{Pr}})$$

$$\lambda_1, \lambda_2 = \sigma\left(h_t \mathrm{W}_h \oplus (I'\mathrm{W}_I + G'_{\mathrm{Pr}}\mathrm{W}_G + W'_{\mathrm{Pr}}\mathrm{W}_W)\right)$$

$$y_t \sim p_t = \mathrm{softmax}(\mathrm{FFN}(h'_t)\mathrm{W}_p + \mathrm{b}_p)$$

where x_t denotes the embedding of current input word; y_t denotes the current target word; σ and $\oplus$ denote the sigmoid function and the matrix-vector addition, respectively; ADA denotes the Adaptive Distilling Attention; The λ_1 and λ_2 weight the importance of G'_{Pr} and W'_{Pr} for each target word, respectively.

3. Experiments

Quantitative Results

Dataset	Methods	Year	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	CIDEr
MIMIC-CXR	CNN-RNN	2015	0.299	0.184	0.121	0.084	0.124	0.263	-
	AdaAtt	2017	0.299	0.185	0.124	0.088	0.118	0.266	-
	Att2in	2017	0.325	0.203	0.136	0.096	0.134	0.276	-
	Up-Down	2018	0.317	0.195	0.130	0.092	0.128	0.267	-
	Transformer	2020	0.314	0.192	0.127	0.090	0.125	0.265	-
	R2Gen	2020	0.353	0.218	0.145	0.103	0.142	0.277	-
	PPKED	Ours	**0.360**	**0.224**	**0.149**	**0.106**	**0.149**	**0.284**	**0.237**
IU-Xray	HRNN	2017	0.439	0.281	0.190	0.133	-	0.342	0.261
	CoAtt	2018	0.455	0.288	0.205	0.154	-	0.369	0.277
	HRGR-Agent	2018	0.438	0.298	0.208	0.151	-	0.322	0.343
	CMAS-RL	2019	0.464	0.301	0.210	0.154	-	0.362	0.275
	Transformer	2020	0.396	0.254	0.179	0.135	0.164	0.342	-
	R2Gen	2020	0.470	0.304	0.219	0.165	0.187	0.371	-
	PPKED	Ours	**0.483**	**0.315**	**0.224**	**0.168**	**0.190**	**0.376**	**0.351**

Table 1. Results of the PPKED and other methods on MIMIC-CXR [1] and IU-Xray [2] datasets.

[1] MIMIC-CXR: A large publicly available database of labeled chest radiographs. arXiv preprint arXiv:1901.07042, 2019.
[2] Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association, 23(2):304–310, 2016.

Qualitative Results

Figure 1. Two examples of ground truth reports and reports generated by HRNN [1] and our method. The Red colored text indicates the abnormalities. The HRNN fails to depict some rare but important abnormalities and generates some error sentences (Blue colored text) and repeated sentences (Underlined text). Our PPKED has higher rate of accurately describing the rare and diverse abnormalities.

Ground Truth:
Lungs are clear. No pleural effusions or pneumothoraces. Heart and mediastinum of normal size and contour. [1]scoliosis.

HRNN:
Heart size is normal. There is a moderate right sided pneumothorax with tip in the right atrium. There is a moderate right sided pneumothorax with large pleural effusion. No pneumothorax masses. No pneumothorax masses. No acute bony abnormalities.

Ours:
[1]There is a scoliosis. No acute cardiopulmonary abnormality. There is no pleural effusion. No evidence of pneumothorax. The lungs are clear. There is no focal airspace consolidation.

Ground Truth:
[1]The heart size is enlarged. [2]The aorta is tortuous. The pulmonary vasculature appears normal. Lungs are otherwise clear bilaterally. No pleural effusions or pneumothorax. No bony abnormalities.

HRNN:
[1]Cardiomegaly with pulmonary vascular congestion and interstitial edema. There is a moderate right sided pneumothorax with large pleural effusion. No bony abnormalities. There is no pneumothorax. There is no pneumothorax.

Ours:
[1]Heart size is enlarged. [2]Tortuosity of the aorta. No pleural effusion. There is no focal airspace consolidation. There is no pneumothorax. No bony abnormalities.

[1] A hierarchical approach for generating descriptive image paragraphs. In CVPR, 2017.

4. Conclusions

Conclusions

- In this work, we present an effective approach of exploring and distilling posterior and prior knowledge for radiology report generation.
- Our approach imitates the working patterns of radiologists to alleviate the data bias problem.
- The experiments demonstrate the effectiveness of our method.
- Our approach not only generates meaningful and robust radiology reports supported with accurate abnormal descriptions and regions, but also outperforms previous state-of-the-art models on the two public datasets.

☺

Thank you for your attention!

If you have any questions about our paper, you can send an email to fenglinliu98@pku.edu.cn