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

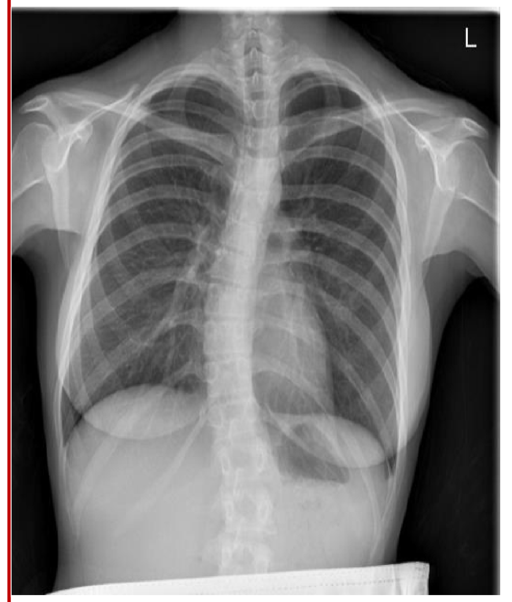
[2] Tencent Medical AI Lab, Beijing, China [3] Peng Cheng Laboratory, Shenzhen, China

* Corresponding authors

PEKING UNIVERSITY

Introduction

- Automatically generating radiology reports systems, which target to produce long and coherent descriptions of medical images, can assist radiologists in clinical decision-making and reduce their workload.



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.

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 Blue colored text stan-ds for the similar sentences used to describe the normalities in ground truths. There are notable data bias and the HRNN fails to depict some rare but important abnormalities and generates some error sentences (Underlined text) and repeated sentences (*Italic* text).

Limitation & Challenge:

- Visual data deviation: the appearance of normal images dominate the data set over that of abnormal images [2].
- Textual data deviation: as shown in Figure 1, in a report, radiologists tend to describe all the items in an image, making the descriptions of normal regions dominate the entire report. Besides, many similar sentences are used to describe the same normal regions.
- The unbalanced visual and textual distributions would distract the model from accurately capturing and describing the rare and diverse abnormalities. As shown in Figure 1, the HRNN [1] generates some repeated sentences of normalities and fails to depict some rare but important abnormalities.

Approach

Solution:

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

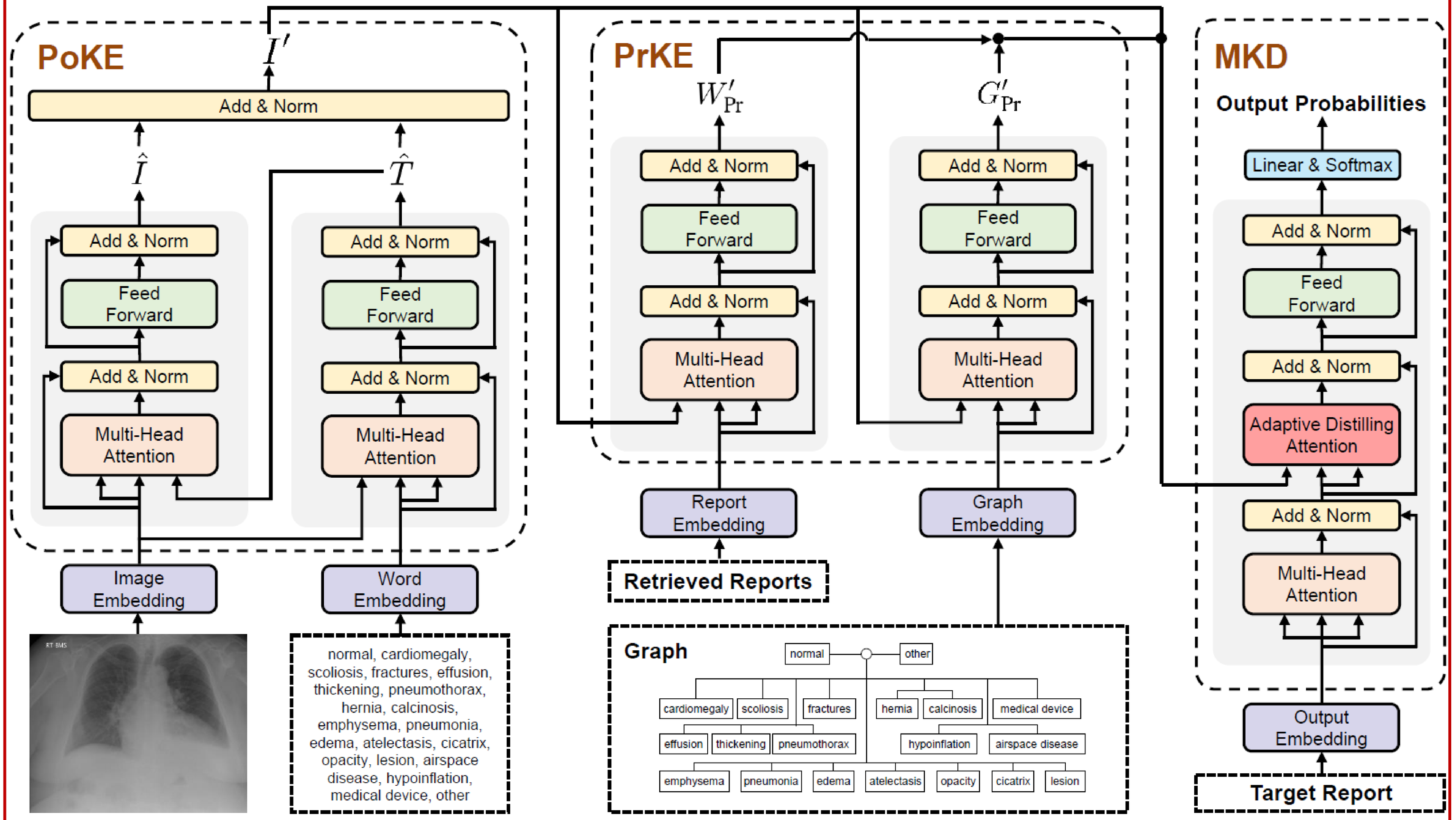


Figure 2. In order to model above working patterns, the PPKED introduces three modules, i.e., **Posterior Knowledge Explorer** (PoKE), **Prio Knowledge Explorer** (PrKE) and **Multi-domain Knowledge Distiller** (MKD).

- Our approach based on the Multi-Head Attention (MHA) and Feed-Forward Network (FFN) from Transformer [3].

Posterior Knowledge Explorer (PoKE): It could alleviate visual data deviation by extracting the abnormal regions based on the input image. 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 $\hat{I}$.

Approach

Prior Knowledge Explorer (PrKE): The PrKE could alleviate textual data deviation by encoding the prior knowledge, including 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:

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

- 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.

Multi-domain Knowledge Distiller (MKD): Finally, the MKD distills the useful knowledge to generate proper reports. Given the embedding of current input word x_t:

$$h_t = \mathrm{MHA}(x_t, x_{1:t}) \qquad \mathrm{ADA}(h_t, I', G'_{\mathrm{Pr}}, W'_{\mathrm{Pr}}) = \mathrm{MHA}(h_t, I' + \lambda_1 G'_{\mathrm{Pr}} + \lambda_2 W'_{\mathrm{Pr}})$$

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

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

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; The λ_1 and λ_2 weight the importance of G'_{Pr} and W'_{Pr} for each target word, respectively.

Experiments

Table 1. Results of the PPKED and other methods on MIMIC-CXR [4] and IU-Xray [5]

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**

- As shown in Table 1, our PPKED outperforms state-of-the-art methods across all metrics on both MIMIC-CXR and IU-Xray datasets.
- As shown in Figure 1, our PPKED has higher rate of accurately describing the rare and diverse abnormalities.

References

[1] A hierarchical approach for generating descriptive image paragraphs. In CVPR, 2017
[2] Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. In CVPR, 2016.
[3] Attention is all you need. In NIPS, 2017.
[4] MIMIC-CXR: A large publicly available database of labeled chest radiographs. arXiv preprint arXiv:1901.07042, 2019.
[5] Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association, 23(2):304–310, 2016.