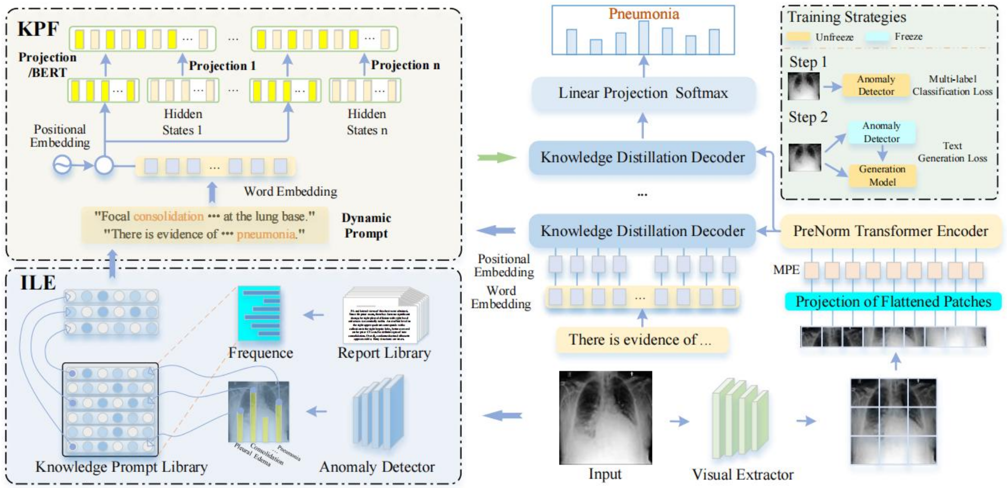
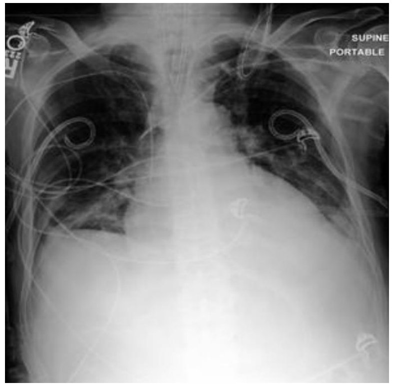
Dynamic Knowledge Prompt for Chest X-ray Report Generation

Shenshen Bu, Yujie Song, Taiji Li, Zhiming Dai
bushsh/songyj9/litj5@mail2.sysu.edu.cn
daizhim@mail.sysu.edu.cn
School of Computer Science and Engineering
Sun Yat-Sen University



1 Background



As compared to the previous radiograph, there is unchanged evidence of a small left apical pneumothorax. On the right, no pneumothorax is seen. The monitoring and support devices, including the bilateral pigtail catheters in the pleural space are unchanged. Minimal increase in bilateral areas of atelectasis at lower lung volumes. No other newly appeared parenchymal opacities. Unchanged moderate cardiomegaly.

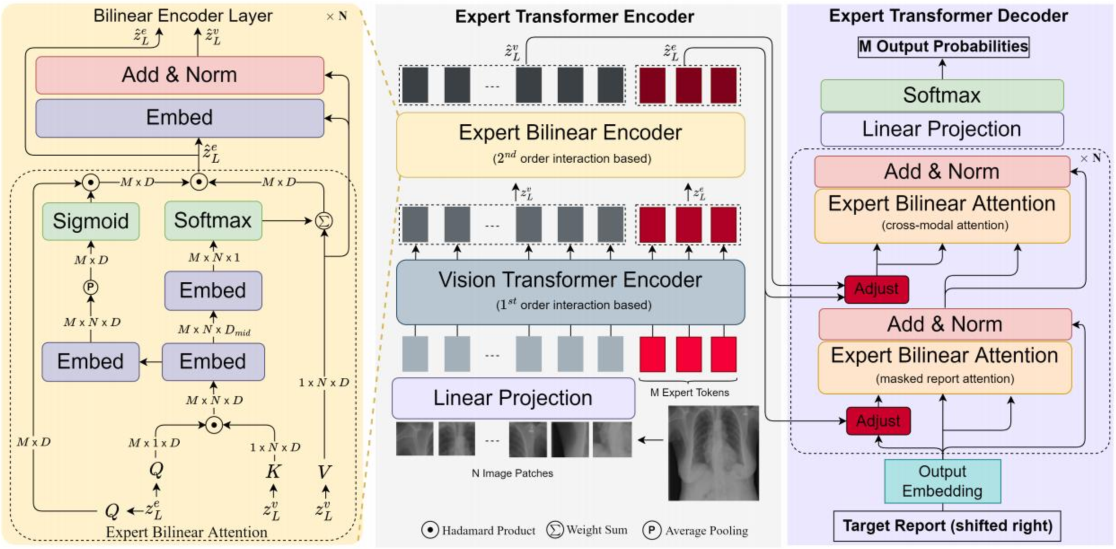
Two challenge：

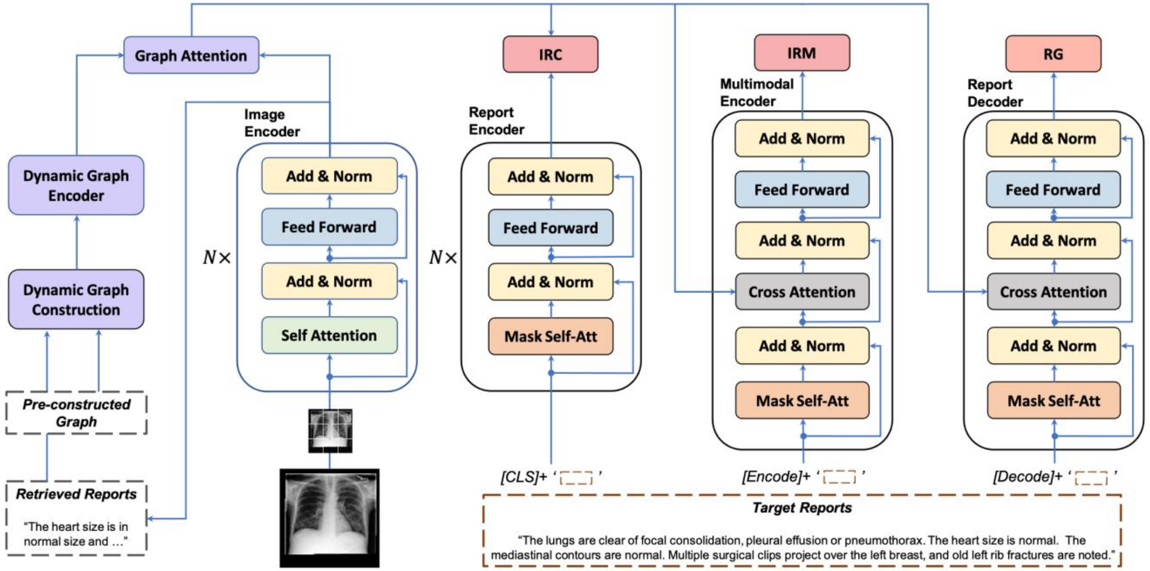
- **Data bias:** Control samples dominate the whole dataset, and the abnormal regions are much smaller than the normal regions in the images of patient samples. Thus, most approaches can learn normal descriptions, but fail to capture anomalies;
- **Visual feature sparsity:** Different from natural images, radiology images lack sufficient discriminative features, which leads to most methods not learning their complex structure and diversity.

1 Background

(a) METransformer

(b) DCL



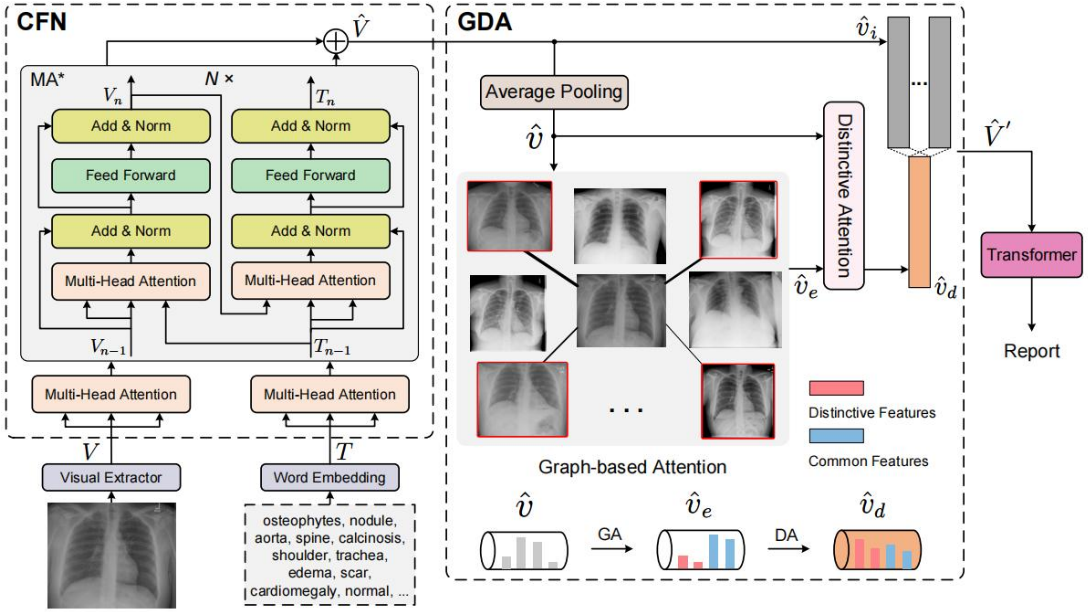


Radiology Report Generation by Transformer with Multiple Learnable Expert Tokens CVPR 2023

Dynamic Graph Enhanced Contrastive Learning for Chest X-ray Report Generation CVPR 2023

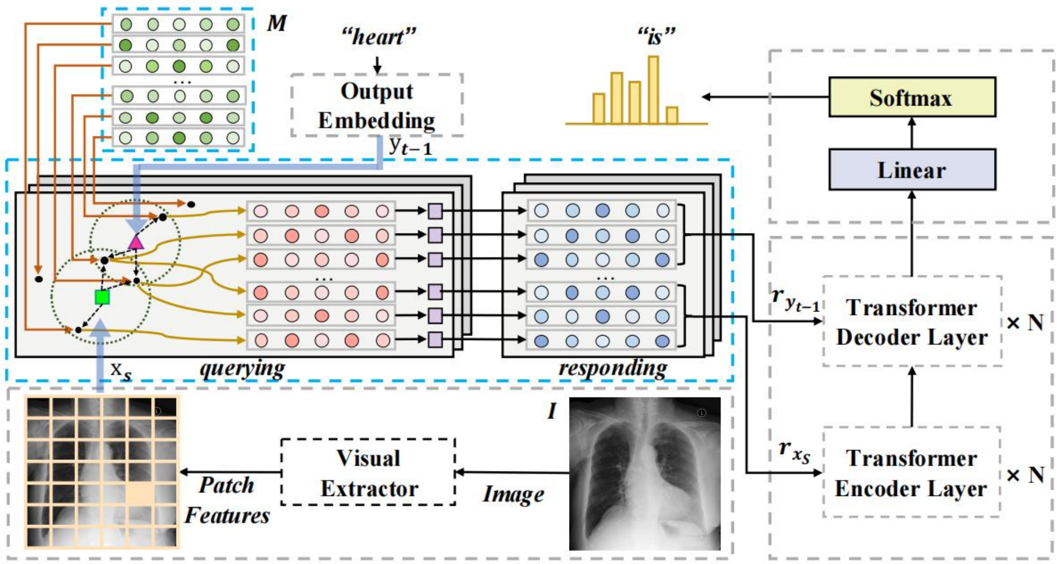
1 Background

(c) TRGD



Embracing Uniqueness: Generating Radiology Reports via a Transformer with Graph-based Distinctive Attention BIBM2022

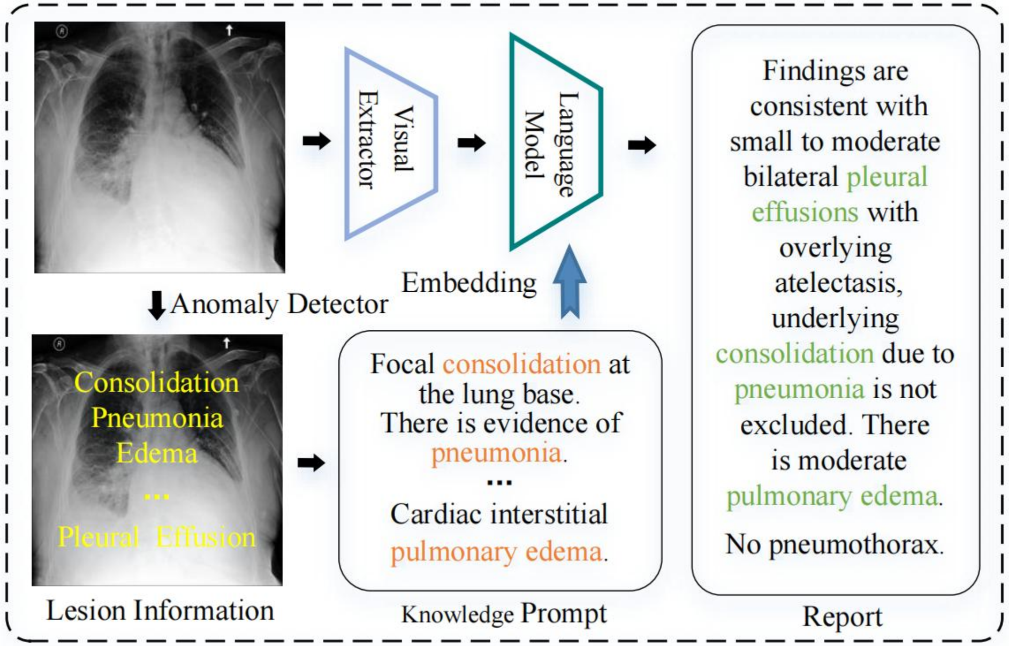
(d) R2GenCMN



Cross-modal Memory Networks for Radiology Report Generation ACL2021

2 Method

2.1 Motivation

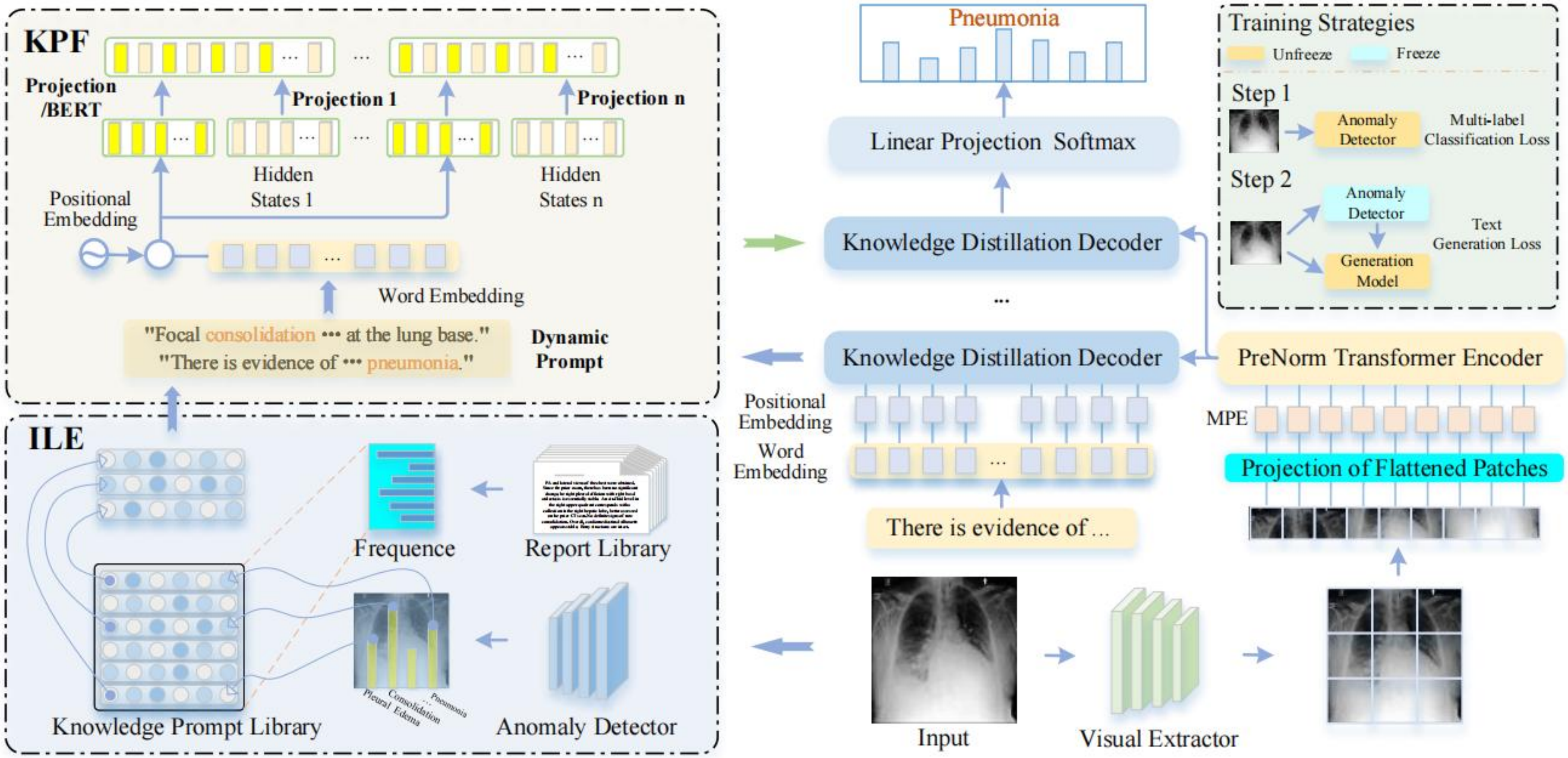


Our proposed DKP generates instance-level knowledge prompt by extracting critical pulmonary lesions information to boost generation.

In radiology image diagnosis, the first step for a radiologist is to focus on abnormal areas and identify lesion locations before creating a comprehensive report. Inspired by this, our method can simulate this process through an anomaly detector that offers precise diagnostic information regarding pulmonary lesions.

2 Method

2.2 Dynamic Knowledge Prompt framework (DKP)



Overview of our DKP architecture. DKP consists of an ILE component for dynamically generating instance-level knowledge prompt, a KPF for generating knowledge-promt-based multi-modal features, a Visual Extractor, a PreNorm Encoder, and a KDD for distilling multi-modal features to generate reports.

2 Method

2.3 Principles for Prompt Library

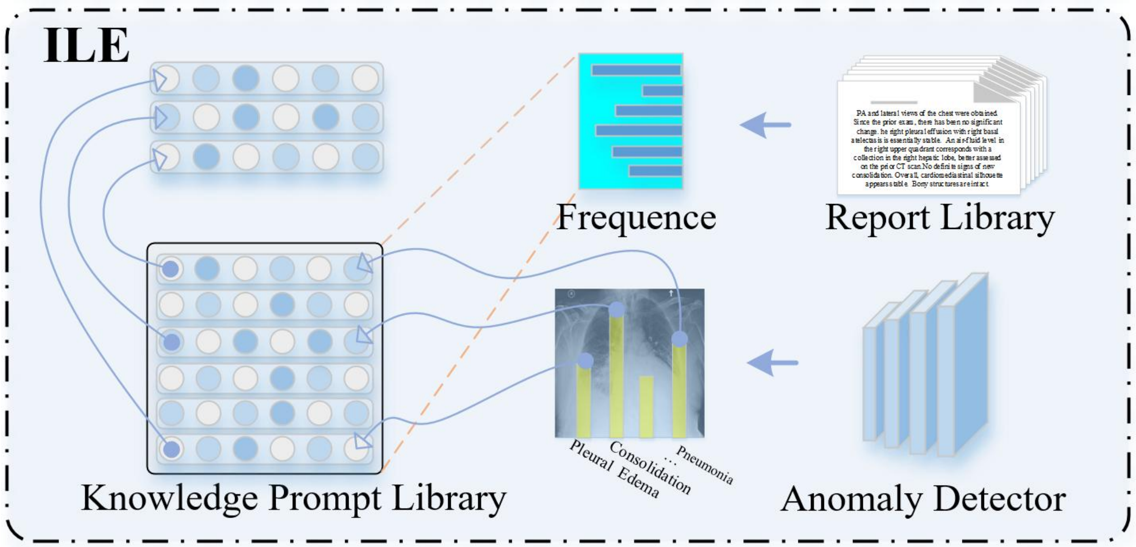
Principles for Prompt Library

"There are bibasilar atelectasis. ",
"There is unchanged evidence of moderate cardiomegaly. ",
"Focal consolidation at the lung base. ",
"Cardiomegaly with moderate interstitial pulmonary edema. ",
"There is evidence of enlarged cardiomediastinum. ",
"There is evidence of bilateral rib fractures. ",
"There is evidence of lung lesion. ",
"Interval development of multiple bilateral airspace opacities. ",
"There is no evidence of focal consolidation... no abnormalities.
","There are bilateral pleural effusions. ",
"There is evidence of pleural. ",
"There is evidence of pneumonia. ",
"There is evidence of pneumothorax. ",
"The monitoring and support devices are constant.

Our approach incorporates two components into the knowledge prompts for each disease, forming the knowledge prompt library. One component utilizes templates such as "the evidence of ..." which aims to enhance natural language generation metrics. The other component includes disease-specific information like "pneumonia" which aims to improve clinical efficacy metrics. By incorporating these components, we address both the issue of smooth readability and the problem of accurate medical diagnosis.

2 Method

2.4 Dynamic Instance Level Explorer (ILE)



Dynamic Instance Level Explorer is utilized to dynamically generate knowledge prompts for each image.

$$p = \sigma(AnomalyDetector(I))$$

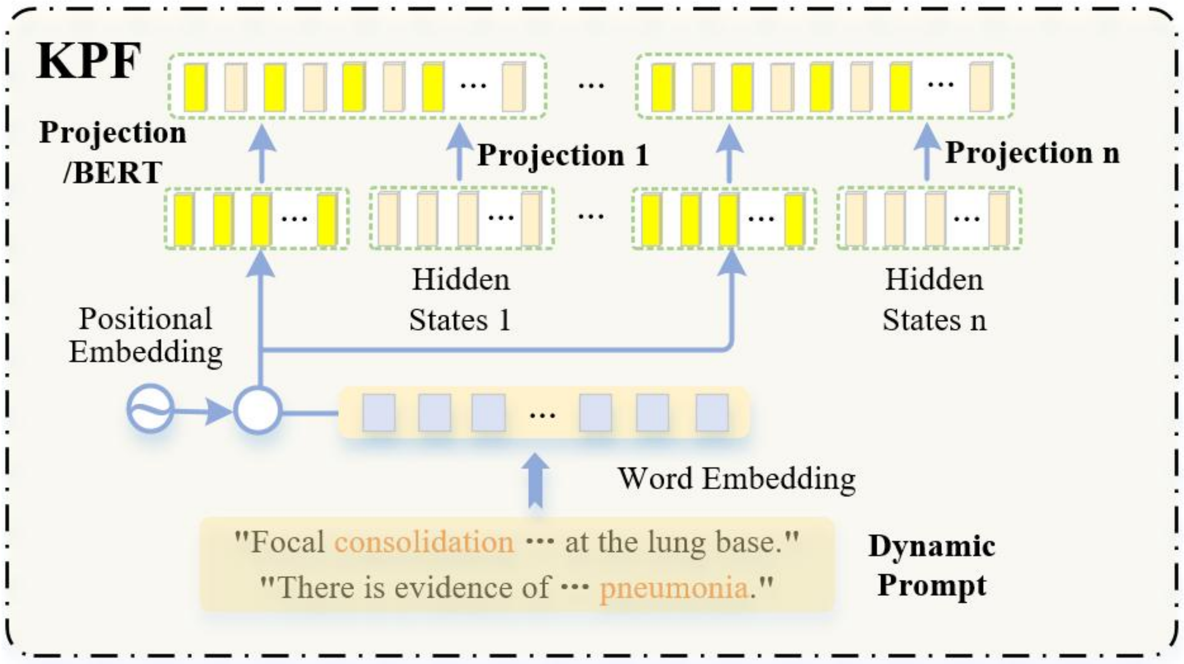
$$Pro(o_n, T) = o_n \odot \{t_1, t_2, ..., t_N\}$$

$$Pro^I = [Pro(o_1, T), ..., Pro(o_N, T)]$$

$$K^I = f_{joint}(f_{sample}(Pro^I))$$

2 Method

2.5 Prior Knowledge Prompt Fuser (KPF)



The KPF is applied to encode the knowledge prompt and merge it with the hidden states generated by the KDD, resulting in multi-modal features.

$$P_e = \{k_1^{idx}, k_2^{idx}, ..., k_L^{idx}\}W^E$$

$$P = P_e + P_{pos}$$

$$K_{hid} = f_h(\{h_1, h_2, ..., h_M\})$$

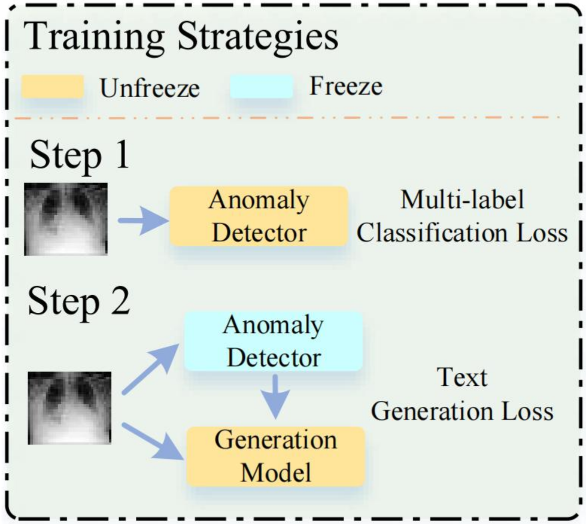
$$K_{pro} = f_p(\{p_1, p_2, ..., p_L\})$$

$$F_{mix} = fusion(K_{hid}, K_{pro})$$

$$K^{mix} = F_{mix}W_{reduce}$$

2 Method

2.6 Two-Stage Training Strategy (OST)



The OST train the ILE exclusively in the first stage and freeze it during the training of language model in the second stage.

Optimization objective for anomaly detector：

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[l_i log(p_i) + (1-l_i) log(1-p_i)]$$

Optimization objective for language generation model：

$$\begin{aligned}\mathcal{L}_{gen}&(X^I, K^I, Y^I; \theta^V, \theta^F, \theta^L)\\ &= CE(f_{\theta^L}(X^I, X^{I\to H}), Y^I)\\ &= -\sum_{i=1}^{M} y_i log(f_{\theta^L(X^I, X^{I\to H})_i})\end{aligned}$$

3 Experimental Setup

Datasets

We evaluate the effectiveness of our DKP on two established benchmarks for report generation: IU X-Ray and MIMIC-CXR：

- IU X-Ray from Indiana University is a collection comprising 7,470 chest X-ray images and 3,955 radiology reports.
- MIMIC-CXR is a extensive chest X-ray dataset curated by Beth Israel Deaconess Medical Center. It includes 473,057 radiographs and 206,563 corresponding reports.

NLG Metrics：

- BLEU
- METEOR
- ROUGE-L

CE Metrics：

- AUROC

4 Experimental Result

4.1 Comparison with State-of-the-Art Methods

DATA	MODEL	Pub.	NLG METRICS BL-1	BL-2	BL-3	BL-4	MTOR	RG	CE METRICS P	R	F1
IU X-Ray	R2Gen	EMNLP20	0.470	0.304	0.219	0.165	0.187	0.371	-	-	-
	PPKED	CVPR21	0.483	0.315	0.224	0.168	-	0.376	-	-	-
	CMCL	ACL21	0.473	0.305	0.217	0.162	0.186	0.378	-	-	-
	R2GenCMN	ACL21	0.475	0.309	0.222	0.170	0.191	0.375	-	-	-
	BLIP	ICML22	0.471	0.294	0.216	0.157	0.186	0.358	-	-	-
	GSKET	MIA22	0.496	0.327	0.238	0.178	-	0.381	-	-	-
	Clinical-BERT	AAAI22	0.495	0.330	0.231	0.170	-	0.376	-	-	-
	DCL	CVPR23	-	-	-	0.163	0.193	0.383	-	-	-
	METransformer	CVPR23	0.483	0.322	0.228	0.172	0.192	0.380	-	-	-
	DKP-BERT	OURS	0.503	0.339	0.241	0.178	0.195	0.392	-	-	-
	DKP-Projection	OURS	**0.507**	**0.344**	**0.245**	**0.181**	**0.214**	**0.398**	-	-	-
MIMIC-CXR	R2Gen	EMNLP20	0.353	0.218	0.145	0.103	0.142	0.277	0.333	0.273	0.276
	PPKED	CVPR21	0.360	0.224	0.149	0.106	0.149	0.284	-	-	-
	CMCL	ACL21	0.344	0.217	0.140	0.097	0.133	0.281	-	-	-
	R2GenCMN	ACL21	0.353	0.218	0.148	0.106	0.142	0.278	0.334	0.275	0.278
	BLIP	ICML22	0.351	0.215	0.146	0.107	0.151	0.265	-	-	-
	GSKET	MIA22	0.363	0.228	0.156	0.115	-	0.284	0.458	0.348	0.371
	Clinical-BERT	AAAI22	0.383	0.230	0.151	0.106	0.144	0.275	0.397	0.435	0.415
	DCL	CVPR23	-	-	-	0.109	0.150	0.284	0.471	0.352	0.373
	METransformer	CVPR23	0.386	0.250	0.169	**0.124**	0.152	**0.291**	0.364	0.309	0.311
	DKP-BERT	OURS	0.412	0.254	0.166	0.115	0.156	0.277	0.487	0.452	0.469
	DKP-Projection	OURS	**0.418**	**0.260**	**0.172**	0.120	**0.159**	0.287	**0.496**	**0.461**	**0.478**

Our proposed DKP is compared with previous state-of-the-art methods on IU X-Ray and MIMIC-CXR datasets. The best scores are in bold face. BL, MTOR and RG refer to BLEU, METEOR and ROUGE, respectively.

4 Experimental Result

4.2 Analysis on Keywords Prediction

	Desc	MODEL			
		BASE	wo/KPF	wo/ILE	DKP
Tis	tube	0.748	0.671	**0.839**	0.711
	atrium	0.664	0.688	0.697	**0.710**
	gastric	0.560	0.620	**0.749**	0.635
	ventricle	0.821	0.766	**0.853**	0.781
Loc	median	0.699	**0.770**	0.686	0.747
	right	0.602	0.641	0.619	**0.657**
	lateral	0.540	0.556	0.537	**0.562**
	above	0.602	**0.668**	0.552	0.634
Ext	enlarged	0.600	0.603	**0.606**	**0.606**
	opacities	0.526	0.539	0.521	**0.560**
	clear	0.618	0.623	0.619	**0.627**
	moderate	0.549	**0.618**	0.571	0.581
Sur	sternotomy	0.745	**0.800**	0.722	0.785
	devices	0.525	0.729	0.539	**0.788**
	pacemaker	0.611	0.703	**0.754**	0.716
	cabg	0.676	**0.734**	0.704	0.732

Comparison AUROC of keywords prediction in MIMIC-CXR. "Tis", "Loc", "Ext", and "Sur" are "tissue", "location", "extent", and "surgery".

	METRICS	MODEL			
		BASE	wo/KPF	wo/ILE	DKP
Macro	Precision	0.412	0.456	0.419	**0.465**
	Recall	0.358	0.430	0.413	**0.446**
	F1-Score	0.352	0.424	0.386	**0.441**
Micro	Precision	0.454	0.491	0.490	**0.506**
	Recall	0.374	0.407	0.405	**0.435**
	F1-Score	0.410	0.445	0.444	**0.468**

The DKP is compared with the prediction results of 16 keywords in MIMIC-CXR.

4 Experimental Result

4.3 Ablation Study

DATA	SETTING	MODEL				NLG METRICS						CE METRICS		
		MPE	KPF	ILE	OST	BL-1	BL-2	BL-3	BL-4	MTOR	RG	P	R	F1
IU X-Ray	BASE				-	0.463	0.287	0.200	0.149	0.178	0.346	-	-	-
	wo/ILE	✓	✓		-	0.489	0.306	0.214	0.156	0.186	0.389	-	-	-
	wo/KPF	✓		✓	-	0.483	0.312	0.224	0.166	0.194	0.390	-	-	-
	wo/MPE		✓	✓	-	0.492	0.319	0.220	0.159	0.207	0.375	-	-	-
	DKP	✓	✓	✓	-	**0.507**	**0.344**	**0.245**	**0.181**	**0.214**	**0.398**	-	-	-
MIMIC -CXR	BASE					0.378	0.223	0.145	0.101	0.140	0.262	0.429	0.348	0.385
	wo/ILE	✓	✓		-	0.395	0.245	0.160	0.112	0.151	0.276	0.487	0.441	0.463
	wo/KPF	✓		✓	✓	0.392	0.238	0.157	0.111	0.150	0.271	0.493	0.420	0.454
	wo/MPE		✓	✓	✓	0.404	0.242	0.155	0.107	0.148	0.272	0.495	0.449	0.472
	wo/OST	✓	✓	✓		0.398	0.245	0.161	0.112	0.158	0.274	0.489	0.443	0.470
	DKP	✓	✓	✓	✓	**0.418**	**0.260**	**0.172**	**0.120**	**0.159**	**0.287**	**0.496**	**0.461**	**0.478**

Ablation experiments of the proposed approach on NLG and CE metrics. The best scores are in bold face and "wo" is defined as "without".

4 Experimental Result

4.4 Case Study

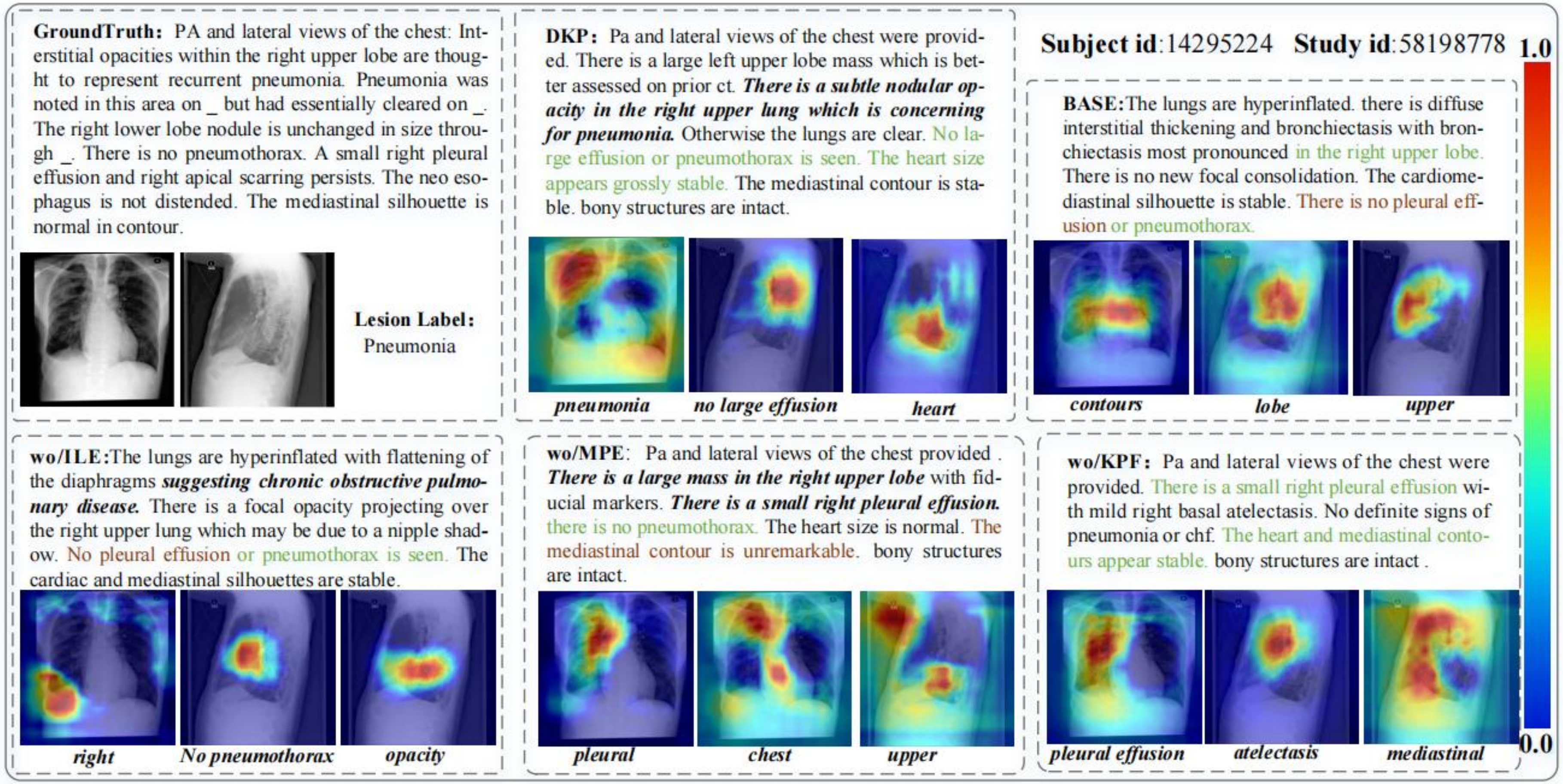


Image-text attention visualizations and case-specific descriptive results from DKP and other baselines. Bold italics and green fonts indicate correct descriptions of the lesion and normal regions, respectively. The red font indicates the error description.

5 Conclusion

In this paper, to alleviate the data bias and visual feature sparse issues, we propose a novel prompt learning paradigm for report generation, which can dynamically generate instance-level knowledge prompts for different cases to boost generation. Extensive experiments and analysis on MIMIC-CXR and IU X-Ray datasets verify the effectiveness of our method. Concretely, DKP effectively improves the quality of radiology report generation by incorporating dynamic knowledge prompt and achieves state-of-the-art performance on both datasets.

Thank You!