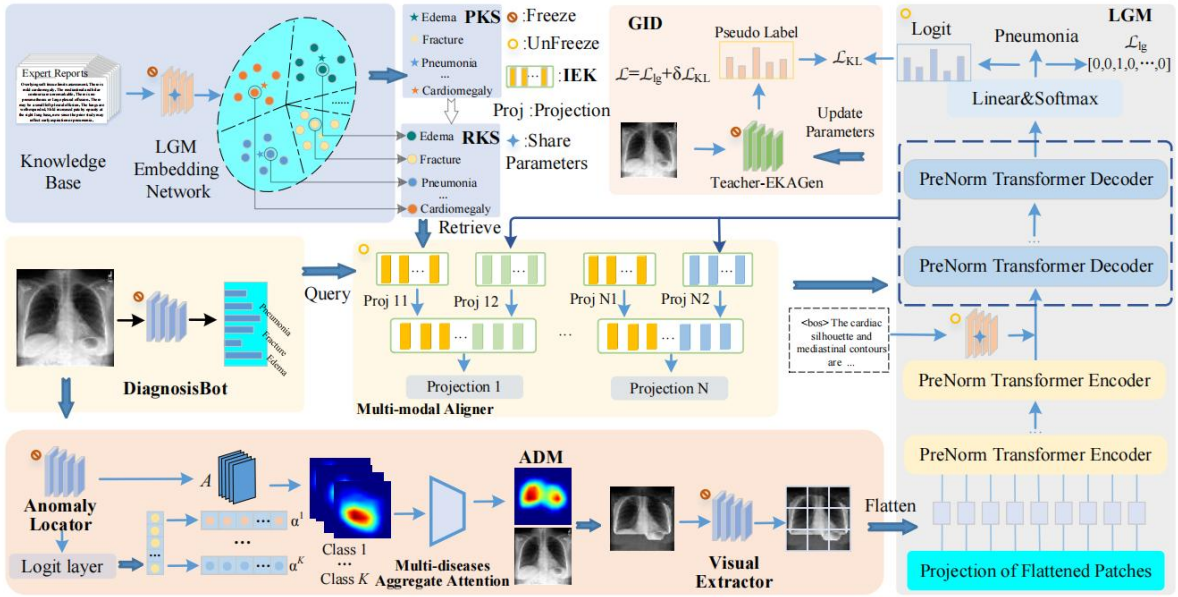
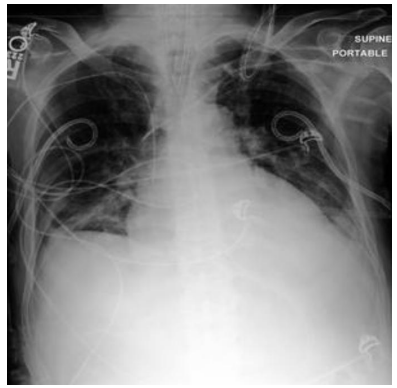
Instance-level Expert Knowledge and Aggregate Discriminative Attention for Radiology Report Generation

Shenshen Bu, Taiji Li, Yuedong Yang, Zhiming Dai
bushsh/litj5@mail2.sysu.edu.cn
yangyd25/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 challenges：

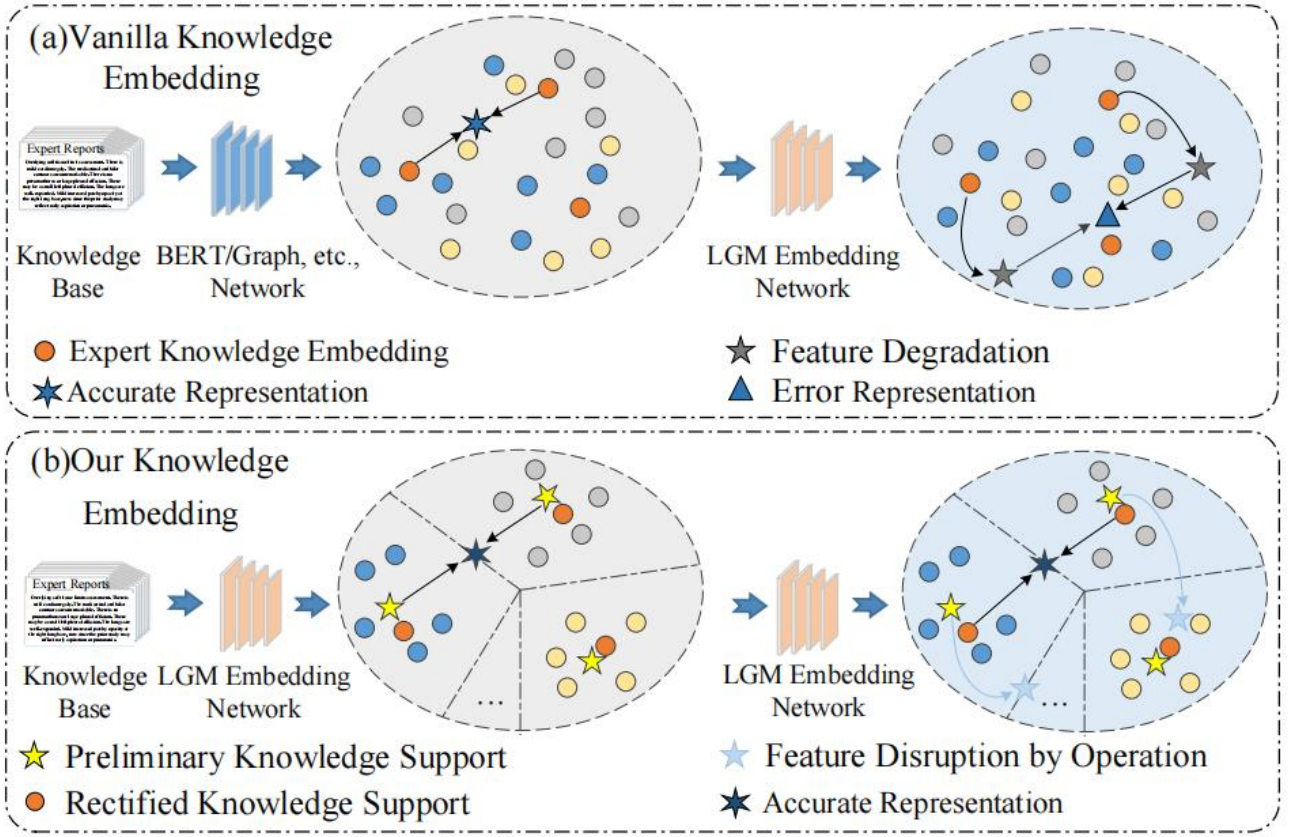
- **Lack discriminative features:** Radiology images often lack discriminative features, which results in a scarcity of reference information for the report generation models.
- **Data deviation:** There is significant data deviation in these datasets due to the rarity of certain diseases, making it challenging to collect positive samples.

2 Method

2.1 Motivations

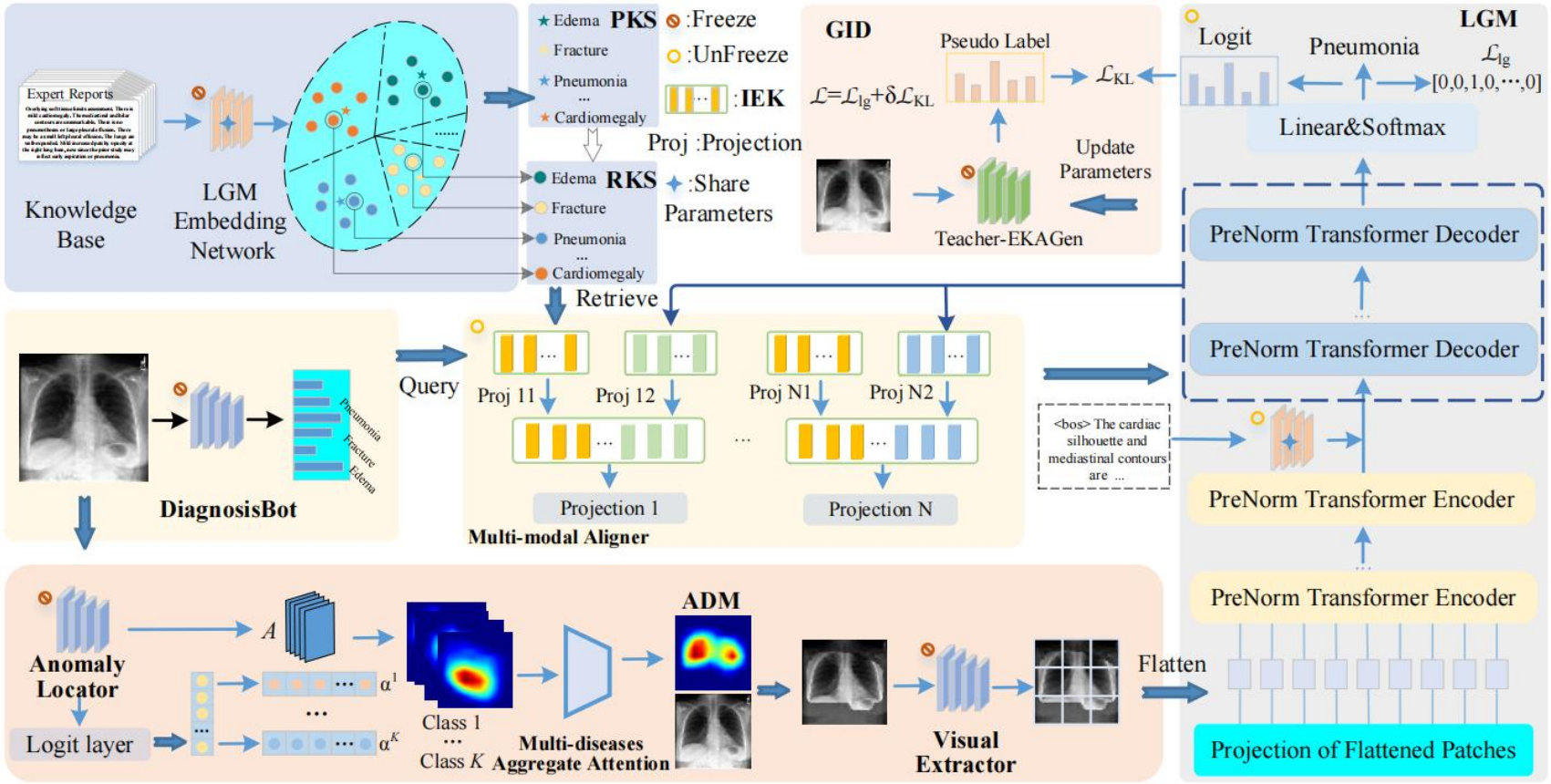
- **Unified embedding network:** Most methods often utilize separate networks to encode prior knowledge, such as BERT and Graph, which leads to inconsistency between these networks and the language generation model (LGM), resulting in feature degradation.

- **Enhance pivotal regions:** Most previous methods fail to enhance attention on pivotal regions of the radiology image, thereby presenting limitations.



2 Method

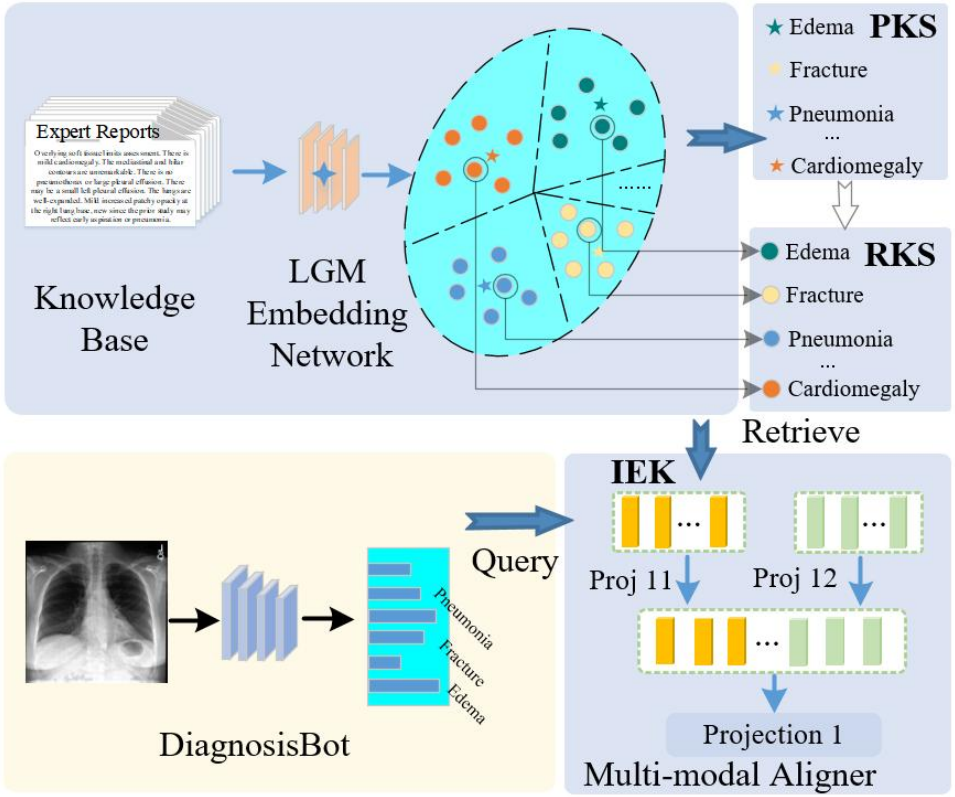
2.2 Instance-level Expert Knowledge and Aggregate Discriminative Attention framework (EKAGen)



EKAGen consists of four components: Instance-level Expert Knowledge (IEK), Aggregate Discriminative Attention Map (ADM), Global Information Self-Distillation (GID), and Language Generation Model (LGM). EKAGen utilizes IEK to address the problem of feature degradation, employs ADM to prioritize pivotal regions, and incorporates GID to distill global knowledge.

2 Method

2.3 Instance-level Expert Knowledge (IEK)



For a given query image, EKAGen utilizes a disease diagnosis module to identify the disease corresponding to the query image and retrieves the corresponding knowledge representation from RKS. Then, it combines all the disease knowledge representations corresponding to the query image to create instance-level expert knowledge (IEK).

$$\mathcal{P}_c = \frac{1}{|\mathcal{X}_c|} \sum_{x_i \in \mathcal{X}_c} \mathcal{F}(x_i)$$

$$\mathcal{K}_c = \underset{\mathcal{F}(x_i)}{\arg\max} \frac{\mathcal{P}_c^T \mathcal{F}(x_i)}{||\mathcal{P}_c|| \cdot ||\mathcal{F}(x_i)||}, x_i \in \mathcal{X}_c$$

$$logit = DiagnosisBot(I)$$

$$\mathcal{C}_i = \begin{cases} \mathbf{E}, & \text{if } \sigma(logit_i) > thre_i \\ \mathbf{0}, & otherwise \end{cases}$$

$$\mathcal{K}^I = concat(\mathcal{K}_1\mathcal{C}_1, ..., \mathcal{K}_i\mathcal{C}_i)$$

2 Method

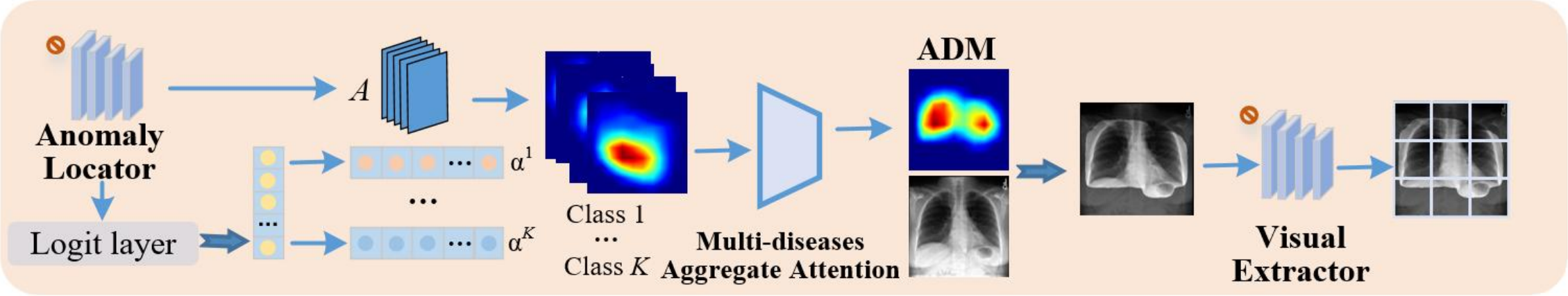
2.4 Aggregate Discriminative Attention Map (ADM)

$$M^I = \sum_{c=1}^{K} ReLU \sum_{k} \alpha_k^c A_{ij}^k$$

$$\mathcal{A}_{ij}^I = \max_{(i',j')\in S} \mathcal{M}_{i+i',j+j'}^I$$

$$\mathcal{M}_{i,j}^I = \begin{cases} 1, & \text{if } M_{i,j} > \theta \\ 0, & otherwise \end{cases}$$

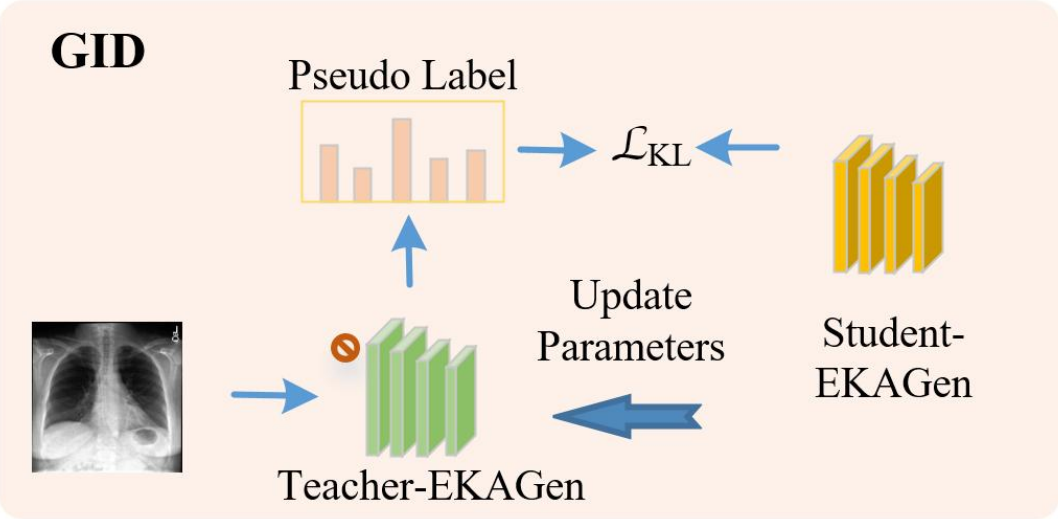
$$Img_{aug} = \mathcal{A}^I \odot I + \gamma(1 - \mathcal{A}^I) \odot I$$



The Aggregate Discriminative Attention Map (ADM) generates discriminative regions in a weakly supervised manner through image-level multi-class labels. It then weakens the signal strength of the background in the input image to constrain the report generation method to focus more on the key areas of radiological images, thereby improving the quality of the generated reports.

2 Method

2.5 Global Information Self-Distillation (GID)

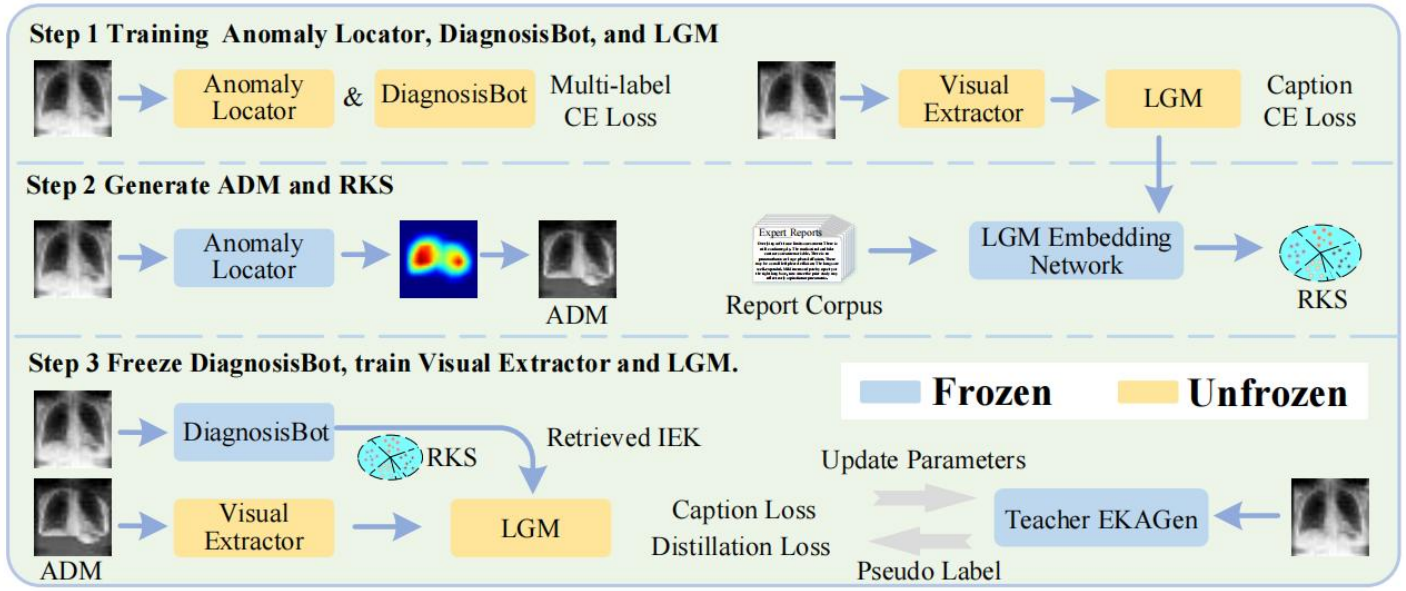


To prevent potential feature erosion, we employ a pre-trained model with a structure identical to EKAGen as a teacher network, taking the entire image as input to distill knowledge into the EKAGen model.

$$\mathcal{L}_{KL} = \frac{1}{N}\sum_{c=1}^{N} KL[p_t(c, I)||p_s(c, I)]$$

2 Method

2.6 Implementation Details of EKAGen



The Implementation of our EKAGen consists of 3 stages.

Optimization objective for DiagnosisBot and Anomaly Locator:

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

Optimization objective for LGM:

$$\mathcal{L}_{lg} = -\sum_{t=1}^{n} logp(y_t^* | y_{<t}^*, \mathcal{K}^{\mathcal{I}}, I)$$

Total optimization objective:

$$\mathcal{L} = \mathcal{L}_{lg} + \delta \mathcal{L}_{KL}$$

3 Experimental Setup

Datasets

We evaluate the effectiveness of our EKAGen 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

Type	Model	IU X-Ray						MIMIC-CXR					
		BL-1	BL-2	BL-3	BL-4	MTOR	RG	BL-1	BL-2	BL-3	BL-4	MTOR	RG
Image Captioning	M2transformer	0.463	0.318	0.214	0.155	-	0.335	0.212	0.128	0.083	0.058	-	0.240
	Grounded	0.446	0.301	0.237	0.176	-	0.343	0.271	0.174	0.122	0.094	-	0.257
Contrastive Based	CA	0.492	0.314	0.222	0.169	0.193	0.381	0.350	0.219	0.152	0.109	0.151	0.283
	DCL	-	-	-	0.163	0.193	0.383	-	-	-	0.109	0.150	0.284
Memory Driven	R2GenCMN	0.475	0.309	0.222	0.170	0.191	0.375	0.353	0.218	0.148	0.106	0.142	0.278
	R2GenRL	0.494	0.321	0.235	0.181	0.201	0.384	0.381	0.232	0.155	0.109	0.151	0.287
Pre Training	BLIP	0.471	0.294	0.216	0.157	0.186	0.358	0.351	0.215	0.146	0.107	0.151	0.265
	Clinical-BERT	0.495	0.330	0.231	0.170	-	0.376	0.383	0.230	0.151	0.106	0.144	0.275
Knowledge Based	GSKET	0.496	0.327	0.238	0.178	-	0.381	0.363	0.228	0.156	0.115	-	0.284
	PPKED	0.483	0.315	0.224	0.168	-	0.376	0.360	0.224	0.149	0.106	0.149	0.284
	KiUT	0.525	0.360	0.251	0.185	**0.242**	**0.409**	0.393	0.243	0.159	0.113	**0.160**	0.285
	METransformer	0.483	0.322	0.228	0.172	0.192	0.380	0.386	0.250	0.169	**0.124**	0.152	**0.291**
Ours	EKAGen (ViT-B/16)	0.517	0.351	0.258	0.191	0.211	**0.409**	0.415	0.254	0.166	0.117	0.154	0.285
	EKAGen (RN-101)	**0.526**	**0.361**	**0.267**	**0.203**	0.214	0.404	**0.419**	**0.258**	**0.170**	0.119	0.157	0.287

Comparing the performance of our EKAGen with other state-of-the-art methods on IU X-Ray and MIMIC-CXR datasets. The abbreviations BL, MTOR, and RG correspond to BLEU, METEOR, and ROUGE, respectively.

4 Experimental Result

4.2 Analysis on Clinical Efficacy Metrics and Backbone

MODEL	MIMIC-CXR		
	Precision	Recall	F1-Score
R2GenCMN	0.334	0.275	0.278
GSKET	0.458	0.348	0.371
Clinical-BERT	0.397	0.435	0.415
KiUT	0.371	0.318	0.321
DCL	0.471	0.352	0.373
METransformer	0.364	0.309	0.311
EKAGen (RN-101)	**0.517**	**0.483**	**0.499**

The comparison of the clinical efficacy metrics on MIMIC-CXR dataset, with the highest scores highlighted in bold.

Dataset	Embedding	BL-1	BL-2	MTOR	RG
IU X-Ray	BERT	0.507	0.348	0.211	0.399
	Uniform	**0.526**	**0.361**	**0.214**	**0.404**
MIMIC-CXR	BERT	0.409	0.251	0.153	0.276
	Uniform	**0.419**	**0.258**	**0.157**	**0.287**

Comparing the performance of a unified prior knowledge encoding network with separate BERT encoding in report generation on the IU X-Ray and MIMIC-CXR datasets.

4 Experimental Result

4.3 Ablation Study

DATA	SETTING	IEK			ADM	GID	NLG METRICS					
		PKS	RKS	EKN			BL-1	BL-2	BL-3	BL-4	MTOR	RG
IU X-Ray	BASE						0.463	0.287	0.200	0.149	0.178	0.346
	(a)	✓					0.475	0.307	0.209	0.148	0.199	0.365
	(b)	-	✓				0.494	0.319	0.217	0.156	0.199	0.379
	(c)	-	✓	✓			0.501	0.328	0.230	0.170	0.206	0.386
	(d)	-	✓	✓	✓		0.509	0.349	0.259	0.198	0.212	0.397
	(e)	-	✓	✓	✓	✓	**0.526**	**0.361**	**0.267**	**0.203**	**0.214**	**0.404**

Quantitative analysis of EKAGen on the IU X-Ray dataset. The BASE model comprises of a Feature Extractor and an Encoder-Decoder structure. The abbreviations BL, MTOR, and RG correspond to the metrics BLEU, METEOR, and ROUGE, respectively.

4 Experimental Result

4.4 Case Study

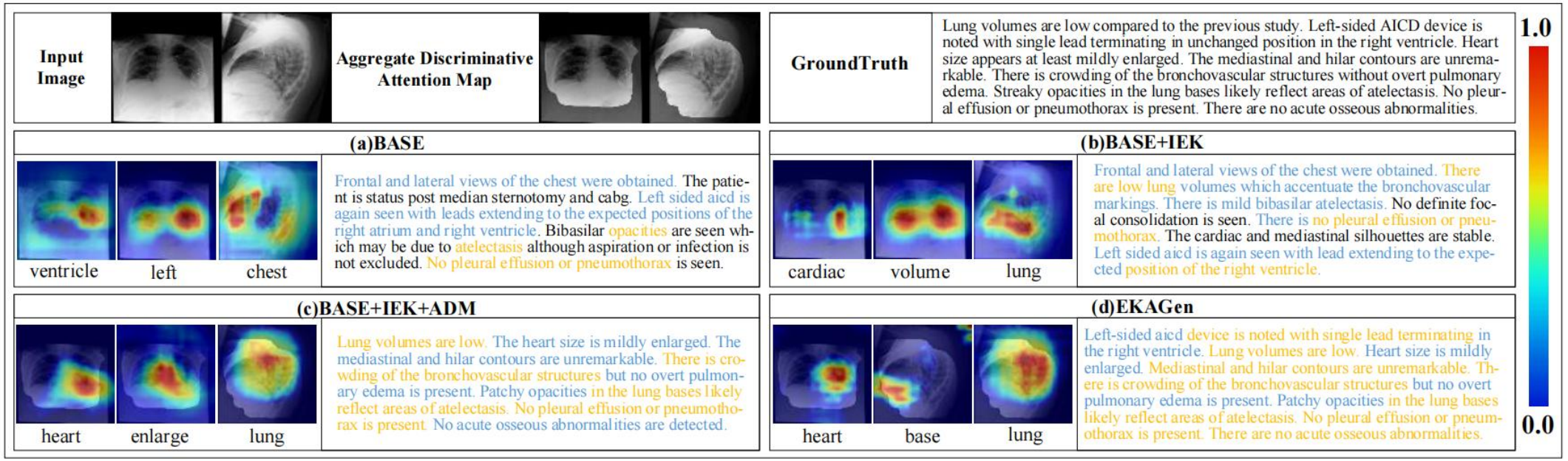


Image-text attention visualizations and captioning results from EKAGen and other models on the MIMIC-CXR dataset. Gold indicates complete alignment with the ground truth, while blue represents semantic alignment.

5 Conclusion

In this paper, we firstly develop comprehensive embedding representations for lung disease and introduce IEK to address the issue of feature degradation. Subsequently, we utilize weak supervision to generate activation maps that highlight critical regions and create ADM to prioritize discriminative regions. Lastly, we propose the GID strategy to prevent feature erosion and provide soft supervision, distilling global knowledge into our model.

Thank You!