UNIVERSITY OF TEXAS ARLINGTON

Journal Club Presentation

Jillur Rahman Saurav

Date: 09-11-2024

Health Data Science Lab
Computer Science & Engineering
University of Texas at Arlington, USA

PromptMRG: Diagnosis-Driven Prompts for Medical Report Generation

Authors: Haibo Jin, Haoxuan Che, Yi Lin, and Hao Chen

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Department of Computer Science and Engineering,
Department of Chemical and Biological Engineering
The Hong Kong University of Science and Technology, Hong Kong, China

DEPARTMENT OF CHEMICAL AND BIOLOGICAL ENGINEERING 化學及生物工程學系

Title: PromptMRG: Diagnosis-Driven Prompts for Medical Report Generation

Authors: Haibo Jin, I

Department of Comp
Department of Chem
The Hong Kong Uni
Kong, China

Categories > Engineering & Computer Science > **Artificial Intelligence**

	Publication	h5-index	h5-median
1.	Neural Information Processing Systems	337	614
2.	International Conference on Learning Representations	304	584
3.	International Conference on Machine Learning	268	424
4.	AAAI Conference on Artificial Intelligence	220	341

Medical Report Generation

Ground-truth: There is opacity to suggest pneumonia.

Medical Report Generation

Ground-truth: There is opacity to suggest pneumonia.

Prediction 1: There is no opacity to suggest pneumonia. ✗

Medical Report Generation

Ground-truth: There is opacity to suggest pneumonia.

Prediction 1: There is no opacity to suggest pneumonia. ✗

Prediction 2: There are right lower lobe opacities concerning for pneumonia. ✓

Medical Report Generation

Ground-truth: There is opacity to suggest pneumonia.

Prediction 1: There is no opacity to suggest pneumonia. ✗

Prediction 2: There are right lower lobe opacities concerning for pneumonia. ✓

Diagnosis-Driven Prompts for Medical Report Generation

Dataset

- MIMIC-CXR
  - Training Set: 270,790
  - Validation Set: 2130
  - Test Set: 3858

- IU X-Ray
  - Test Set: 2955

EXAMINATION: CHEST (PA AND LAT)

INDICATION: ___ year old woman with ?pleural effusion // ?pleural effusion

TECHNIQUE: Chest PA and lateral

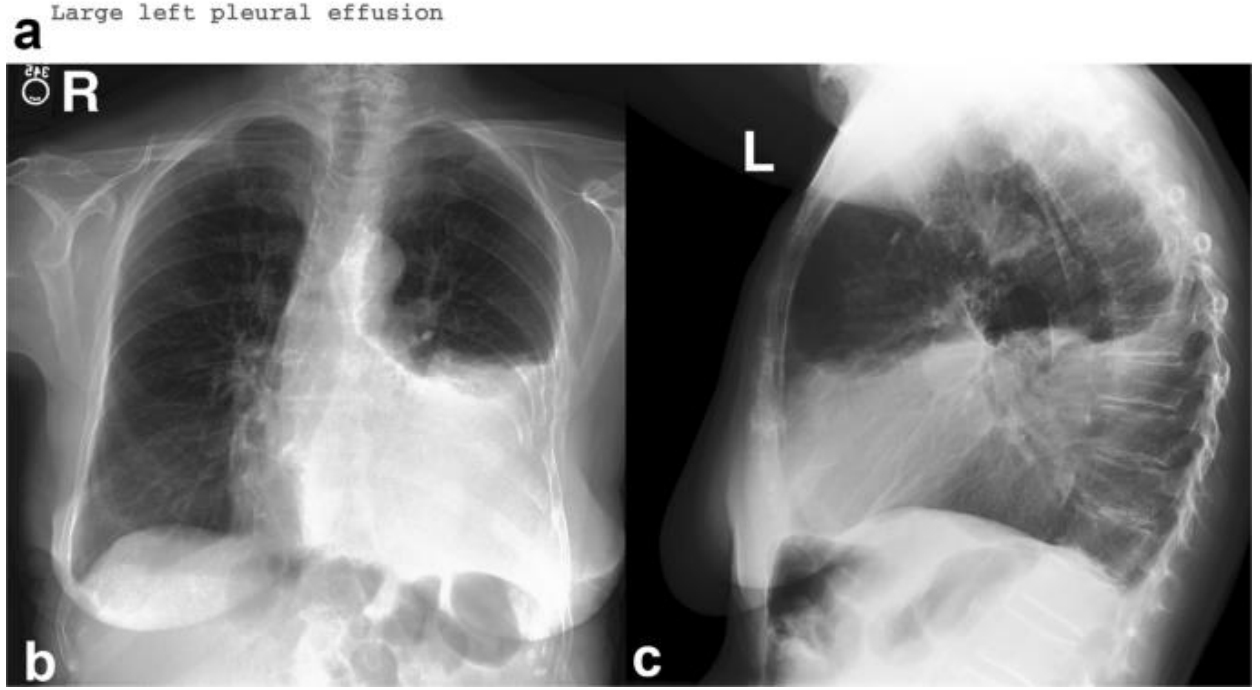
COMPARISON: ___

FINDINGS:

Cardiac size cannot be evaluated. Large left pleural effusion is new. Small right effusion is new. The upper lungs are clear. Right lower lobe opacities are better seen in prior CT. There is no pneumothorax. There are mild degenerative changes in the thoracic spine

IMPRESSION:

Large left pleural effusion



Johnson et al. (2019). MIMIC-CXR: A de-identified chest X-ray database. Scientific Data, 6(1), 317.

Architecture

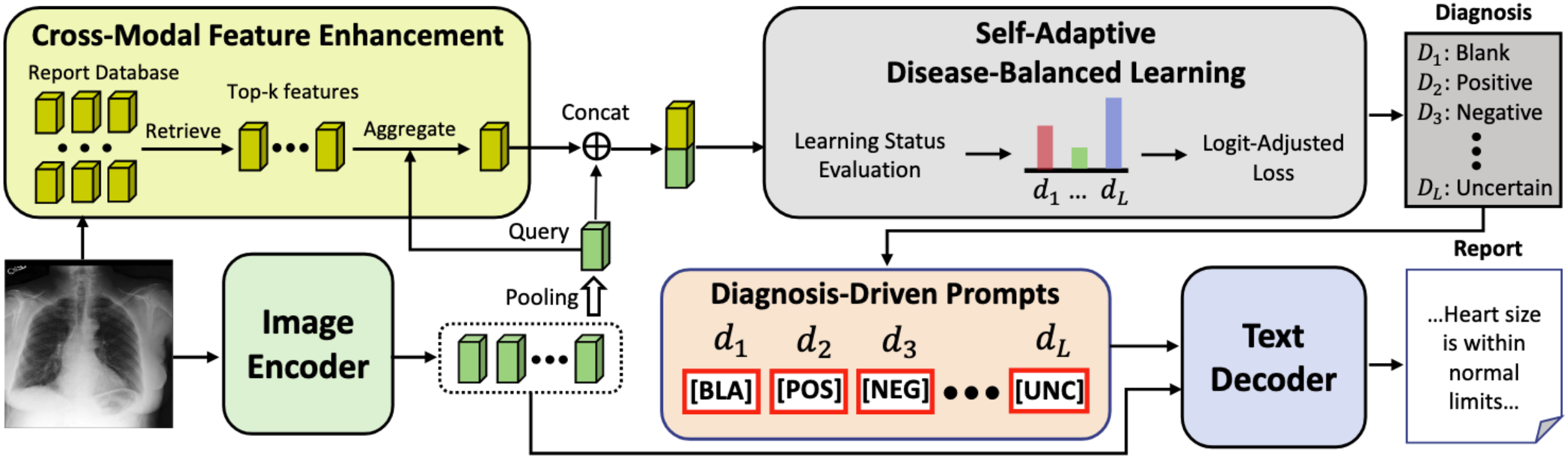
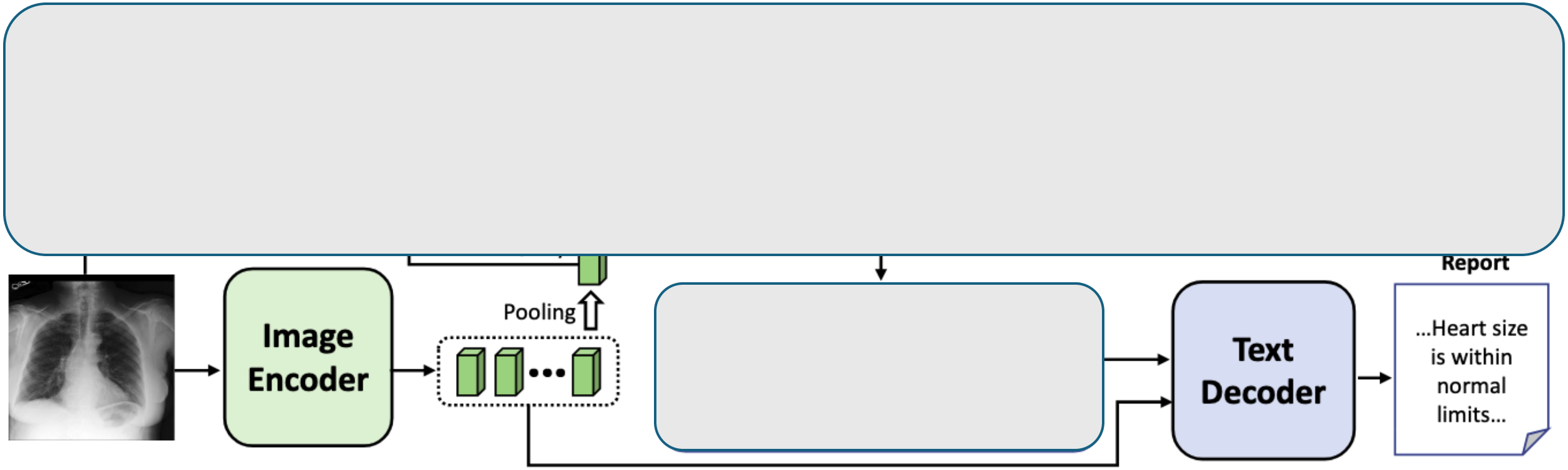


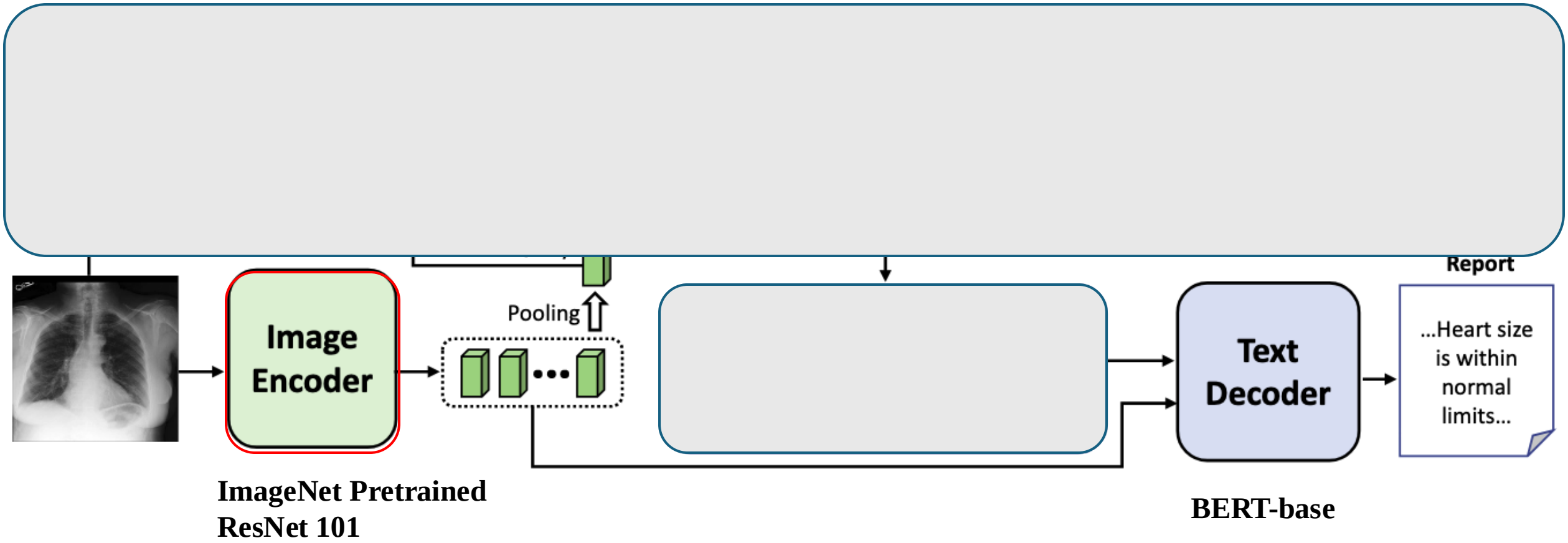
Figure 2: The overall framework of PromptMRG, which mainly consists of an image encoder and a text decoder for report generation. The diagnosis-driven prompts module is proposed to guide the decoder for diagnostically correct reports. The cross-modal feature enhancement is designed to enhance the feature for disease classification via a report database. The self-adaptive disease-balanced learning is further proposed to handle the imbalanced performance among diseases.

Architecture

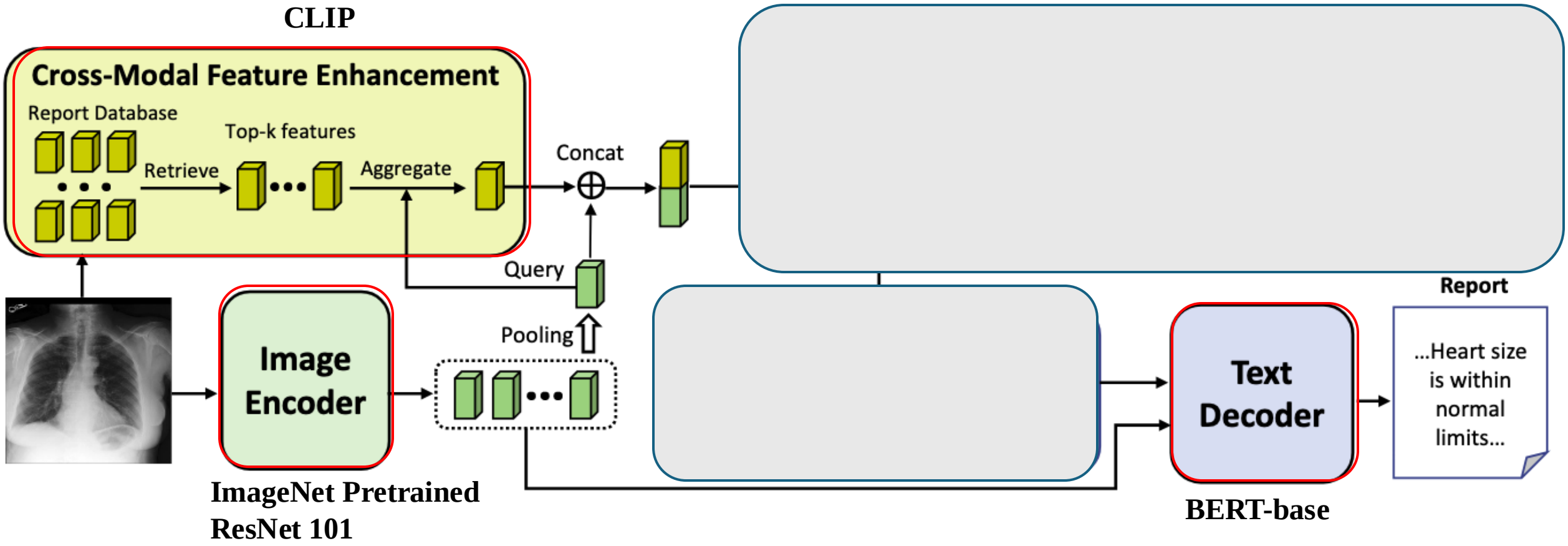


Encoder Decoder Architecture

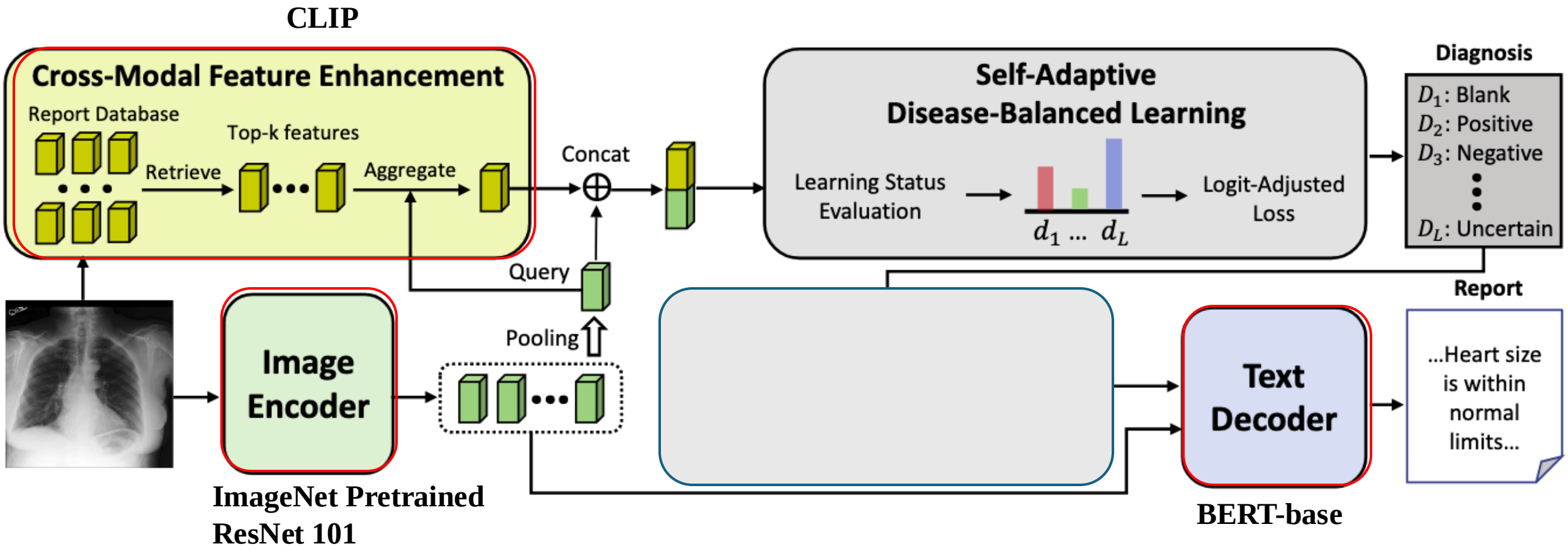
Architecture



Architecture



Architecture



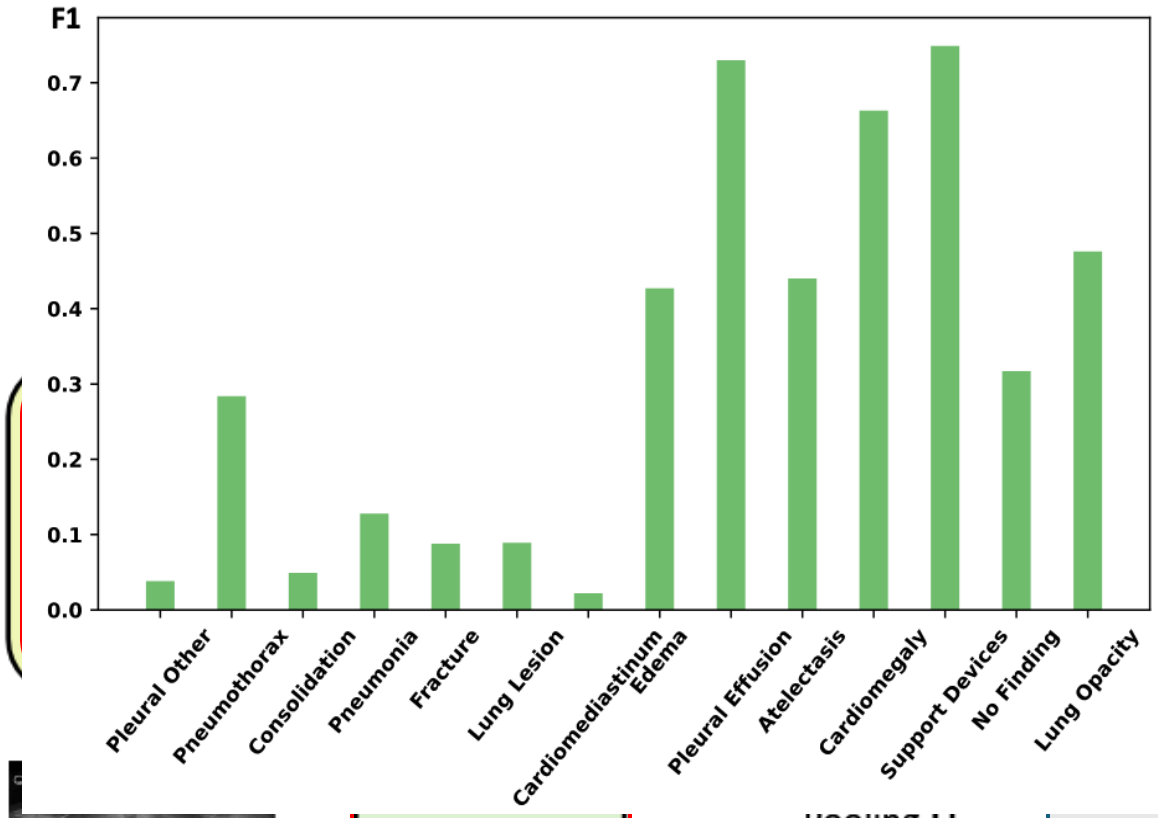


Image Encoder

Pooling

ImageNet Pretrained
ResNet 101

Self-Adaptive
Disease-Balanced Learning

Learning Status
Evaluation

d_1 ... d_L

Logit-Adjusted
Loss

Diagnosis

D_1: Blank
D_2: Positive
D_3: Negative
D_L: Uncertain

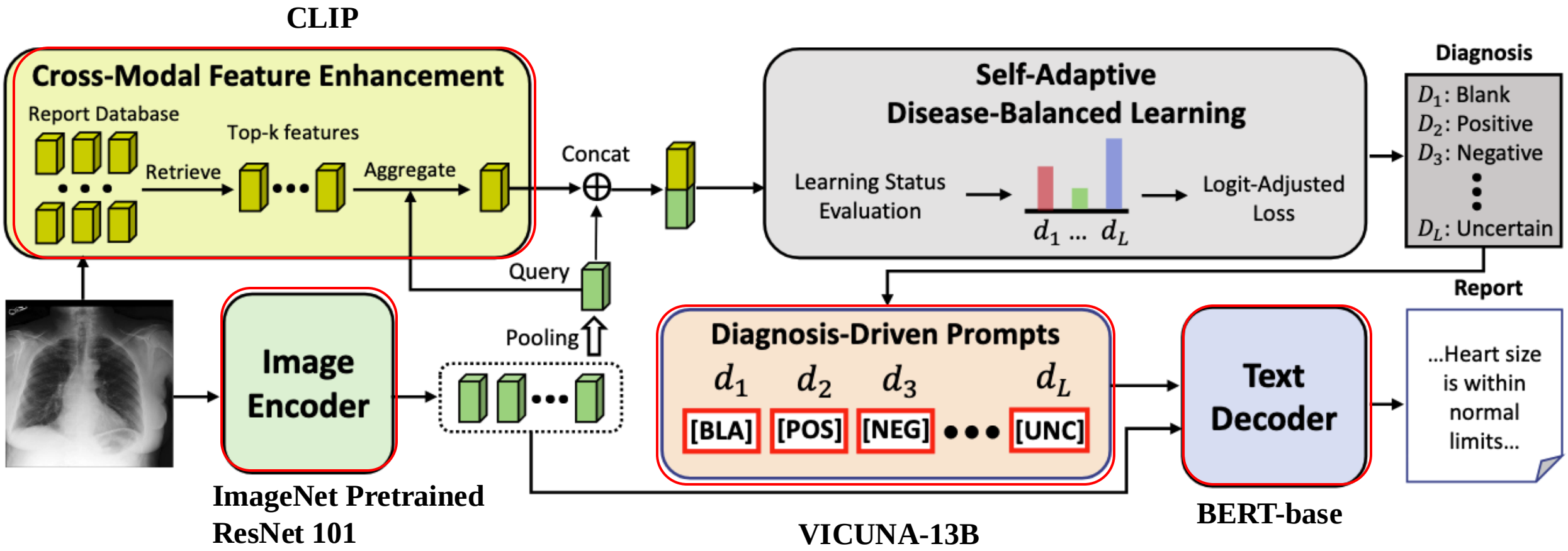
Report

Text
Decoder

BERT-base

...Heart size
is within
normal
limits...

Architecture



Results

Dataset	Model	Year	CE Metrics			NLG Metrics			
			Precision	Recall	F1	BLEU-1	BLEU-4	METEOR	ROUGE-L
MIMIC	R2Gen	2020	0.333	0.273	0.276	0.353	0.103	0.142	0.277
	M2TR	2021	0.240	0.428	0.308	0.378	0.107	0.145	0.272
	MKSG	2022	0.458	0.348	0.371	0.363	0.115	-	0.284
	CliBert	2022	0.397	0.435	0.415	0.383	0.106	0.144	0.275
	CVT2Dis.*	2022	0.356	0.412	0.384	0.392	0.124	0.153	0.285
	M2KT	2023	0.420	0.339	0.352	0.386	0.111	-	0.274
	METrans.	2023	0.364	0.309	0.311	0.386	0.124	0.152	**0.291**
	KiUT	2023	0.371	0.318	0.321	0.393	0.113	0.160	0.285
	DCL	2023	0.471	0.352	0.373	-	0.109	0.150	0.284
	RGRG*#	2023	0.461	0.475	0.447	0.373	**0.126**	**0.168**	0.264
	Ours	-	**0.501**	**0.509**	**0.476**	**0.398**	0.112	0.157	0.268
IU X-Ray	R2Gen†	2020	0.141	0.136	0.136	0.325	0.059	0.131	0.253
	CVT2Dis.*†	2022	0.174	0.172	0.168	0.383	0.082	0.147	0.277
	M2KT†	2023	0.153	0.145	0.145	0.371	0.078	0.153	0.261
	DCL†	2023	0.168	0.167	0.162	0.354	0.074	0.152	0.267
	RGRG*†	2023	0.183	0.187	0.180	0.266	0.063	0.146	0.180
	Ours	-	**0.213**	**0.229**	**0.211**	**0.401**	**0.098**	**0.160**	**0.281**

Table 1: Comparison with SOTA MRG methods on MIMIC-CXR and IU X-Ray. $*$ indicates the used image size is larger than 224. † indicates the performance evaluated by us. # indicates the usage of a different data split. The best results are in **bold**.

Results

DDP	ADL	CFE	SDL	CE Metrics			NLG Metrics			
				Precision	Recall	F1	BLEU-1	BLEU-4	METEOR	ROUGE-L
✘	✘	✘	✘	0.430	0.368	0.370	0.397	**0.116**	**0.158**	**0.272**
✔	✘	✘	✘	0.487	0.461	0.444	0.391	0.106	0.151	0.261
✔	✔	✘	✘	0.496	0.466	0.451	0.393	0.106	0.152	0.261
✔	✔	✔	✘	**0.514**	0.475	0.464	0.394	0.108	0.152	0.263
✔	✔	✘	✔	0.489	0.500	0.468	0.397	0.111	0.155	0.264
✔	✔	✔	✔	0.501	**0.509**	**0.476**	**0.398**	0.112	0.157	0.268

Table 2: Ablation study of each module on MIMIC test set.

Results

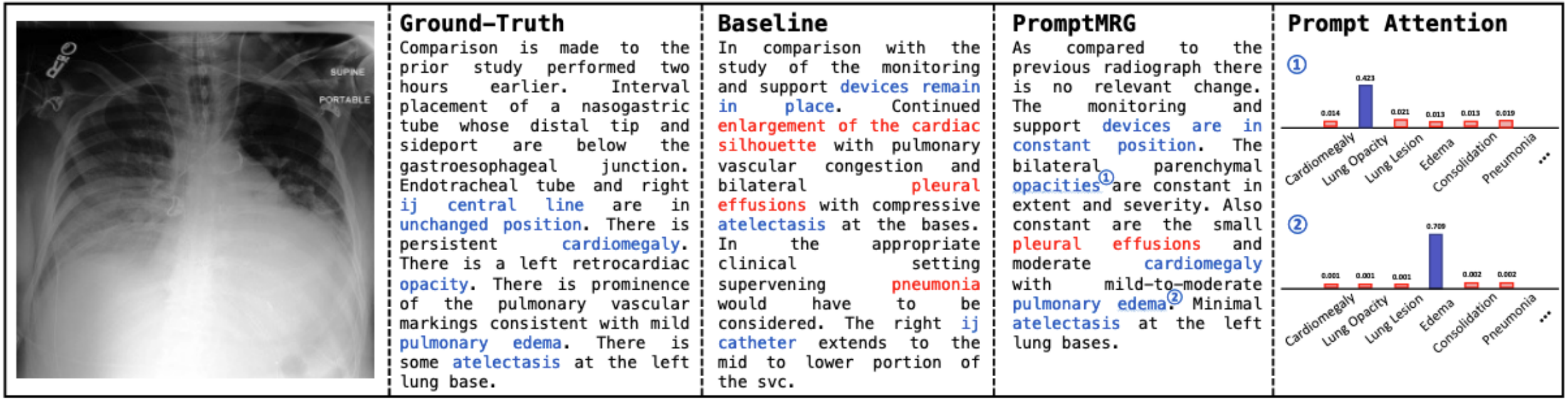


Figure 5: Qualitative examples of the baseline and the proposed method. Blue font indicates consistent content with the ground-truth while red font indicates incorrect content. See more examples in Appendix.

Implementation

- Encoder: ResNet-101 (ImageNet)
- Decoder: Bert-base
- Optimizer: AdamW, weight decay 0.05
- Learning Rate: 5e-5, cosine schedule
- Epochs: 10, Batch Size: 16, Image Size: 224
- Framework: PyTorch 2.0
- Hardware: RTX 3090, ~24 hours

Key Takeaways

- Identify limitations of existing models
- Leverage pretrained models for each modality
- Experiment with prompt engineering
- Perform ablation studies
- Validate with different approaches