Contrastive Attention for Automatic Chest X-ray Report Generation

Fenglin Liu[1*], Changchang Yin[2], Xian Wu[3], Shen Ge[3], Ping Zhang[2], Xu Sun[1]

[1] Peking University, China

[2]The Ohio State University, USA

[3] Tencent Medical AI Lab, China

PEKING UNIVERSITY

Chest X-ray Report Generation

- **Task Definition**: It aims to generate a long paragraph describing both the normal and abnormal regions, which can assist radiologists in clinical decision-making.
- **Task Objectives**:
  - ➢ a long and coherent report.
  - ➢ cover key medical findings:
    - ✓e.g., heart size and lung opacity.
  - ➢ correctly describe any abnormalities and its details:
    - ✓e.g., the location and shape of the abnormality.
  - ➢ correctly describe potential diseases:
    - ✓e.g., effusion and consolidation.

Indication: No acute cardiopulmonary abnormality.

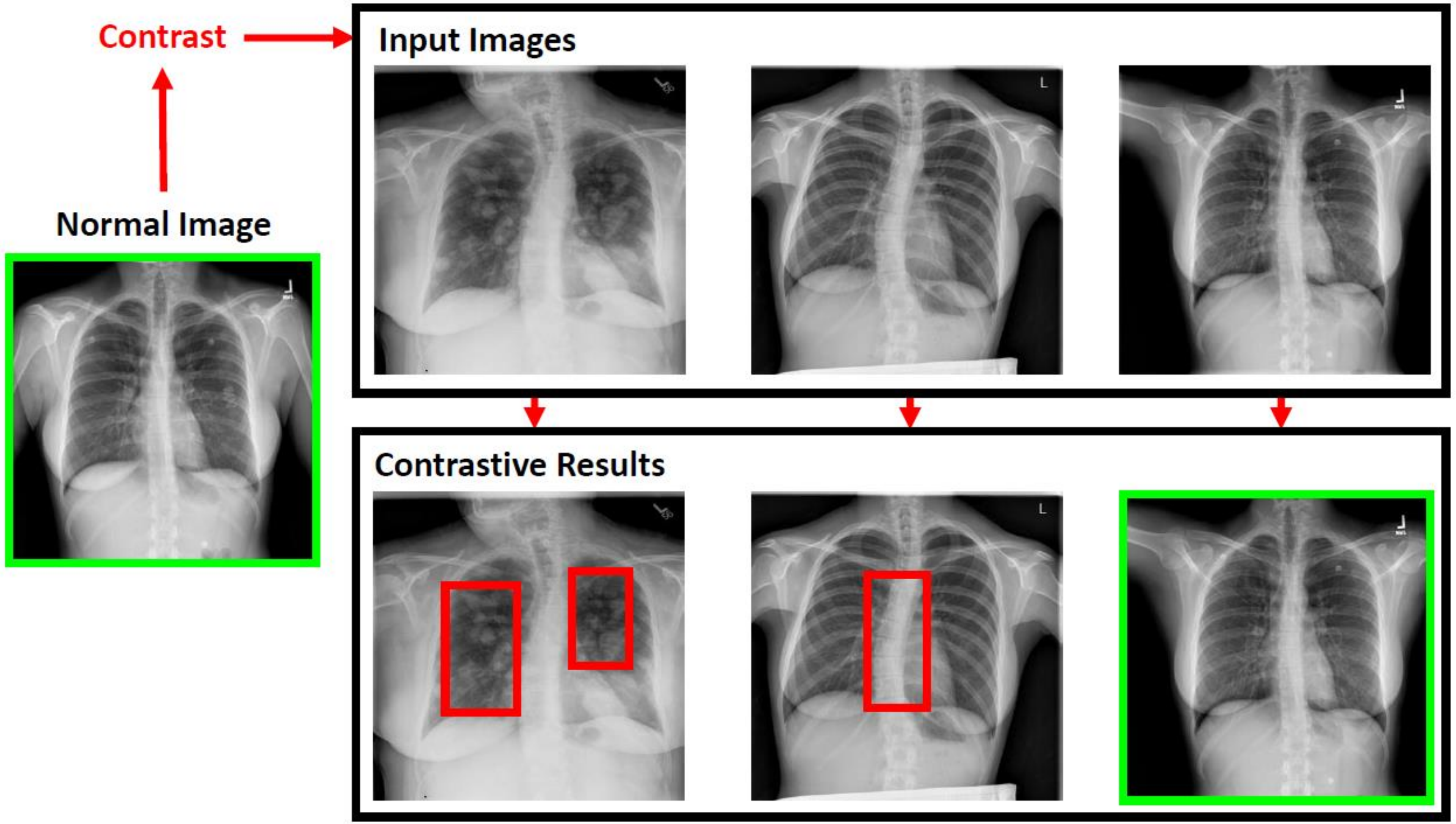
Findings: Lungs are clear without focal infiltrates. Calcified right upper lobe granuloma unchanged from prior. No pneumothorax or pleural effusion. Normal heart size. Normal pulmonary vascularity. Bony thorax intact.

Impression: No acute cardiopulmonary abnormality.

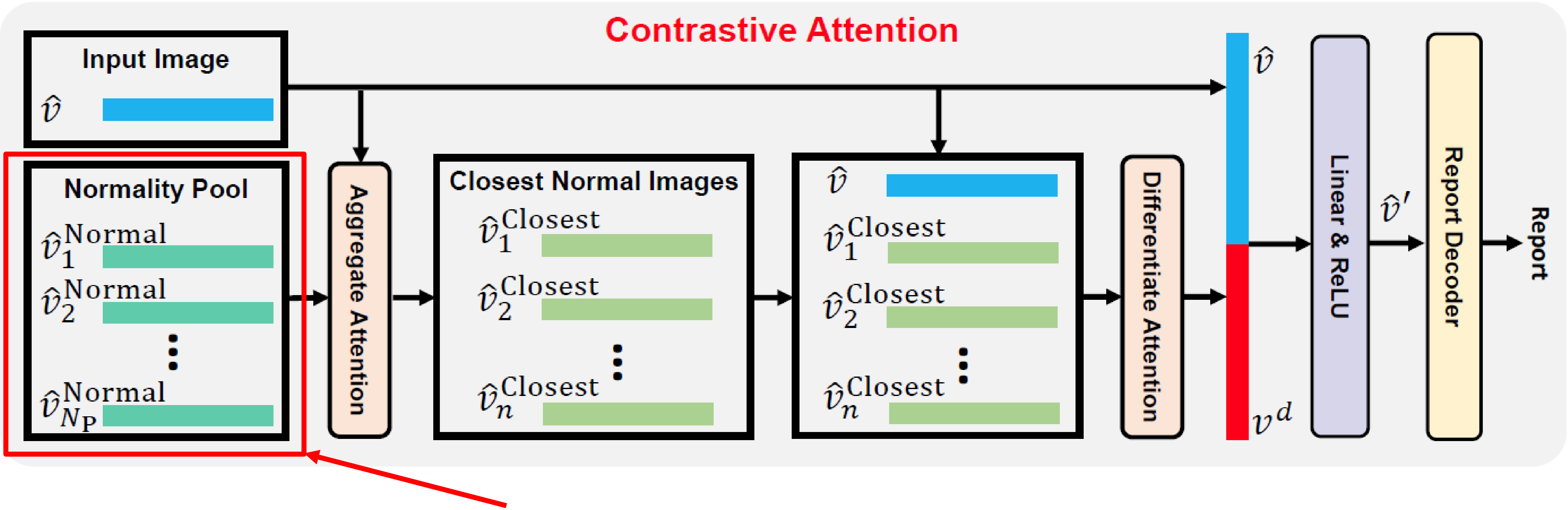
Tags: Calcified Granuloma

Motivation

- **Urgent goal and core value**: correctly **capturing** and **describing** the abnormalities.
- To capture the abnormal regions of given chest X-ray image, a natural intuition is to compare it with normal images and identify the differences. Therefore, we propose the Contrastive Attention to enable existing methods to better capture and describe the abnormalities.

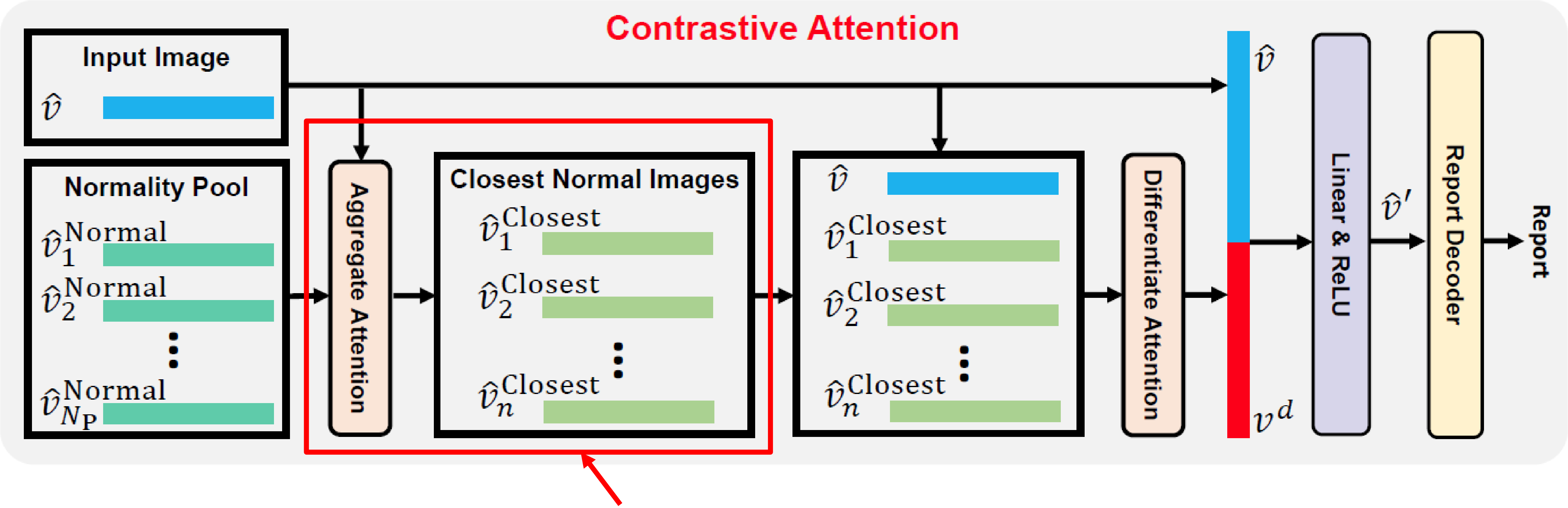


Approach



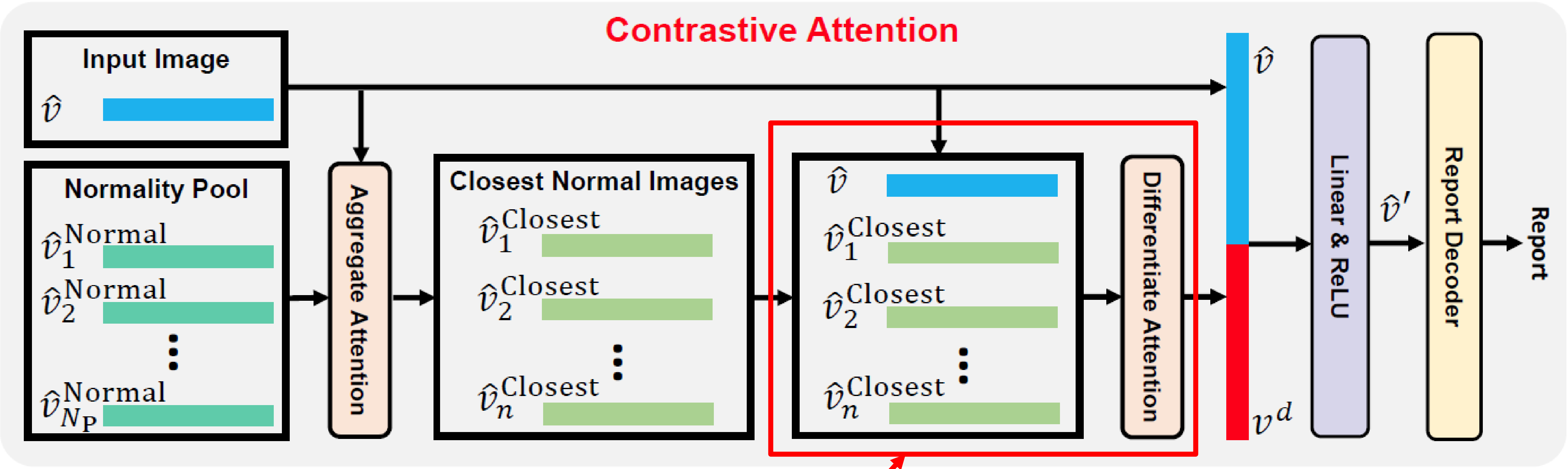
- We first build a set of normal images which are all extracted from the training dataset.

Approach



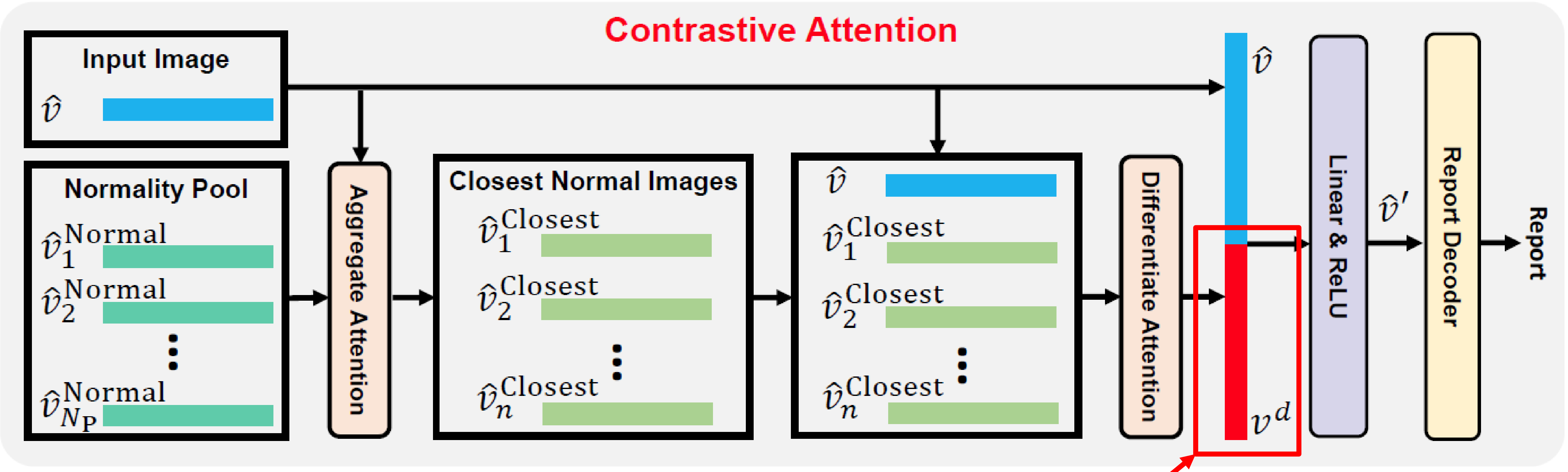
- We introduce the **Aggregate Attention** to prioritize normal images that are closer to the current input image, and filter out normal images that appear differently.

Approach



- We further introduce the **Differentiate Attention** to distill the common features between the input image and the refined normal images.

Approach



- Then, the acquired common features are subtracted from the visual features of the input image. In this manner, the residual visual features of the input image are treated as the contrastive information that captures the differentiating properties between input image and normal images.

Experiments

Settings	Methods	Dataset: MIMIC-CXR (Johnson et al., 2019)						Dataset: IU-X-ray (Demner-Fushman et al., 2016)					
		B-1	B-2	B-3	B-4	M	R-L	B-1	B-2	B-3	B-4	M	R-L
(a)	NIC (Vinyals et al., 2015)†	0.290	0.182	0.119	0.081	0.112	0.249	0.352	0.227	0.154	0.109	0.133	0.313
	w/ Contrastive Attention	0.317	0.200	0.127	0.089	0.120	0.262	0.368	0.232	0.166	0.118	0.144	0.323
(b)	Visual-Attention (Xu et al., 2015)†	0.318	0.186	0.122	0.085	0.119	0.267	0.371	0.233	0.159	0.118	0.147	0.320
	w/ Contrastive Attention	0.309	0.202	0.129	0.093	0.122	0.265	0.384	0.245	0.172	0.125	0.141	0.315
(c)	Spatial-Attention (Lu et al., 2017)†	0.302	0.189	0.122	0.082	0.120	0.259	0.374	0.235	0.158	0.120	0.146	0.322
	w/ Contrastive Attention	0.320	0.204	0.129	0.091	0.122	0.266	0.378	0.236	0.161	0.116	0.146	0.335
(d)	Att2in (Rennie et al., 2017)†	0.314	0.199	0.126	0.087	0.125	0.265	0.410	0.257	0.173	0.131	0.149	0.325
	w/ Contrastive Attention	0.327	0.205	0.132	0.095	0.124	0.271	0.442	0.281	0.200	0.150	0.171	0.344
(e)	Adaptive-Attention (Lu et al., 2017)†	0.307	0.192	0.124	0.084	0.119	0.262	0.433	0.285	0.194	0.137	0.166	0.349
	w/ Contrastive Attention	0.330	0.208	0.134	0.095	0.126	0.270	0.425	0.279	0.198	0.142	0.167	0.347
(f)	Up-Down (Anderson et al., 2018)†	0.318	0.203	0.128	0.089	0.123	0.266	0.389	0.251	0.170	0.126	0.154	0.317
	w/ Contrastive Attention	0.336	0.209	0.134	0.097	0.128	0.273	0.378	0.246	0.169	0.129	0.152	0.330
(g)	HLSTM (Krause et al., 2017)†	0.321	0.203	0.129	0.092	0.125	0.270	0.435	0.280	0.187	0.131	0.173	0.346
	w/ Contrastive Attention	0.352	0.216	0.145	0.105	0.139	0.276	0.453	0.290	0.203	0.153	0.178	0.361
(h)	HLSTM+att+Dual (Harzig et al., 2019)†	0.328	0.204	0.127	0.090	0.122	0.267	0.447	0.289	0.192	0.144	0.175	0.358
	w/ Contrastive Attention	0.323	0.202	0.130	0.102	0.138	0.277	0.464	0.292	0.205	0.149	0.176	0.364
(i)	Co-Attention (Jing et al., 2018)†	0.329	0.206	0.133	0.095	0.129	0.273	0.463	0.293	0.207	0.155	0.178	0.365
	w/ Contrastive Attention	0.351	0.213	0.148	0.106	0.147	0.270	0.486	0.311	0.223	0.178	0.187	0.372
(j)	Multi-Attention (Huang et al., 2019)†	0.337	0.211	0.136	0.097	0.130	0.274	0.468	0.299	0.211	0.155	0.180	0.366
	w/ Contrastive Attention	0.350	0.219	0.152	0.109	0.151	0.283	0.492	0.314	0.222	0.169	0.193	0.381

- Existing methods equipped with our Contrastive Attention outperform baselines.

北京大学 PEKING UNIVERSITY

Conclusions

- In this paper, we propose the Contrastive Attention model to capture abnormal regions by contrasting the input image and normal images for chest X-ray report generation.

- The experiments on two public datasets demonstrate the effectiveness of our approach, which can be easily incorporated into existing models to boost their performance under most metrics.

☺

Thank you for your attention!

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