Supplementary Document: Visual-Textual Attentive Semantic Consistency for Medical Report Generation

Yi Zhou[1], Lei Huang[2], Tao Zhou[3], Huazhu Fu[4], and Ling Shao[4]

[1]School of Computer Science and Engineering, Southeast University, Nanjing, China
[2]SKLSDE, Institute of Artificial Intelligence, Beihang University, Beijing, China
[3]School of Computer Science and Technology, Nanjing University of Science and Technology, China
[4]Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

A. Discussion of disease and description pattern labels.

The disease labels and description pattern labels (DPLs) are extracted from the medical reports using CheXpert Labeler. Someone might wonder if the description pattern labels are associated with the disease labels, or they are not related. In the current version, once the disease and DPLs are extracted, they are treated independently. When we were selecting the DPLs according to the MIMIC-CXR and IU X-ray datasets, we didn't observe many duplicated DPLs in a report. Moreover, since the DPLs contain useful information like lesion location and size, they are very helpful to train the image encoder and MMSA module. However, the concern for more open-world medical reports is insightful, where duplicated DPLs may appear for multiple diseases in a report (maybe severely ill patients). The relationships between disease labels and DPLs could be exploited in the decoder design to generate more accurate sentences.

B. Training if some report parts are missing.

We can consider what happens if the finding part or the indication parts are missing in the original report. Will that affect the training? First, the finding part, which contains detailed sentence descriptions, is usually the main part of a report. This is our target to generate. It is hard to implement a medical report generator without any training report data. Second, if the indication part is missing, we can detach the clinical feature encoding part, which will decrease the performance to some extent. However, the indication part is usually available in reports' raw data from a hospital. Most existing public datasets contain the indication part together with the finding part.

C. Discussion of $\mathcal{L}_{ISM}$ and $\mathcal{L}_{DA}$.

Here we further discuss $\mathcal{L}_{ISM}$ and $\mathcal{L}_{DA}$ with their individual contributions. As shown in Eq. 10 in the original paper, since $\{\mathcal{L}_{sent} + \mathcal{L}_{stop} + \mathcal{L}_{word}\}$ is the basic loss function used in the medical report generation framework, we set it as the baseline(B). We show the effectiveness of $\mathcal{L}_{ISM}$ and $\mathcal{L}_{DA}$ in the following table on MIMIC-CXR, when other modules are fully deployed. E.g. BLEU-1 shows that $B + \mathcal{L}_{ISM}(\sim\ 97.8\%)$ contributes more than $B + \mathcal{L}_{DA}(\sim 94.4\%)$. Although the two losses totally bring around 3% additional training time, there is no extra cost in the testing.

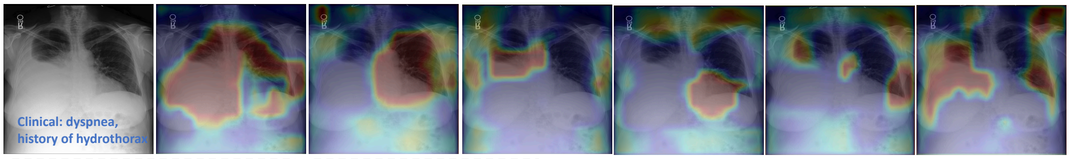
Methods	BLEU-1	CIDEr	ROUGE	METEOR	nKTD
Baseline(B)	0.345	0.898	0.321	0.168	0.169
B+$\mathcal{L}_{ISM}$	0.364	1.029	0.325	0.179	0.132
B+$\mathcal{L}_{DA}$	0.351	0.954	0.320	0.173	0.155
B+$\mathcal{L}_{ISM}$+$\mathcal{L}_{DA}$	0.372	1.121	0.335	0.190	0.106

D. Qualitative Results.

We demonstrate more results showing the reports generated by our method, compared with ground-truths and some baselines in Fig. 1.

E. Acknowledgements.

This work was supported by the National Natural Science Foundation of China (62106043), and the Fundamental Research Funds for the Central Universities, China (2242021k30056). We also thank the collaborated radiologists who helped us to carefully select the description pattern labels and provided many great suggestions for this work.



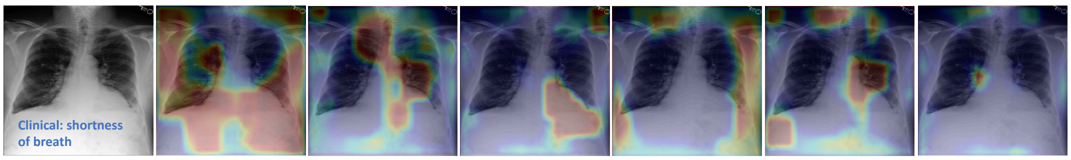
Ours: a large right pleural effusion is noted and the right lung cannot be assessed for consolidation. patchy opacity at the left base and in the left lower lobe may reflect atelectasis. right-sided atelectasis is present. the heart size is normal. no acute osseous abnormality is detected. there is moderate to large right pleural effusion.

Ours-wo-VSMA: there is a right pleural effusion. the cardiac and mediastinal silhouettes are normal. no acute cardiopulmonary abnormalities. there is no focal infiltrate or effusion. the endotracheal tube is in appropriate positions. no pneumonia or pneumothorax.

Ours-wo-DPL: cadiomediastinal silhouette is stable and within normal limits. atelectasis opacities without focal consolidation pneumothorax. there is right pleural effusion. the heart size is normal. no acute cardiopulmonary abnormality. no acute bony abnormality.

Ours-wo- $\mathcal{L}_{ISM}$&$\mathcal{L}_{DA}$: there is a moderate to large right pleural effusion. moderate to severe pulmonary edema and moderate pleural effusion both on the right. there is mild cardiomegaly. left lung is clear without focal consolidation. no overt pulmonary edema is seen. small hiatal hernia.

Ground Truth: pa and lateral views of the chest provided. there is a large right pleural effusion new from prior exam with associated compressive atelectasis in the right middle lower lung. the left lung is clear aside from mild left lower lung atelectasis. heart size difficult to assess given effacement of the right heart border. bony structures are intact. large right pleural effusion with compressive atelectasis in the right middle lower lung.



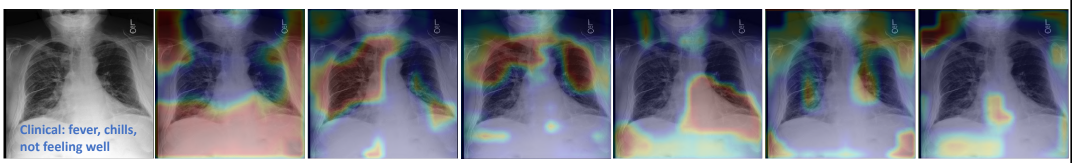
Ours: the finding represent bilateral pleural effusions. the aorta is tortuous and demonstrates atherosclerotic calcifications. cardiomegaly without convincing signs of pneumonia or edema. the visualized osseous structures demonstrate no acute abnormality. prominent on the left with compressive atelectasis at the bases. there is no focal consolidation.

Ours-wo-VSMA: moderate atelectasis is noted on the left and overall lung volumes are low. heart size is likely normal. there is no focal consolidation. no pleural effusion or pneumothorax is seen. no acute cardiopulmonary process. opacity on the left side may be due to pulmonary fluid and atelectasis.

Ours-wo-DPL: there may be a component of a pleural effusion. the heart is enlarged and unchanged. aortic knob remains calcified. there is no focal consolidation. retrocardiac opacities most likely represent atelectasis and pneumonia can not be excluded. no pneumothorax is present.

Ours-wo- $\mathcal{L}_{ISM}$&$\mathcal{L}_{DA}$: no signs of acute or chronic parenchymal infiltrates is present. heart size and cardiomediastinal contours are normal. there is no consolidation or pleural effusion. no pneumonia or effusion is noted. no acute osseous abnormality is identified. there is no identified displaced rib fracture. normal chest radiograph.

Ground Truth: mild enlargement of the cardiac silhouette is unchanged. the aorta is diffusely calcified. no pulmonary edema is present and the hilar contours are normal. small bilateral pleural effusions are likely unchanged with persistent patchy atelectasis at the lung bases more so on the left. no pneumothorax is present. no acute osseous abnormality is detected. small bilateral pleural effusions and bibasilar atelectasis.



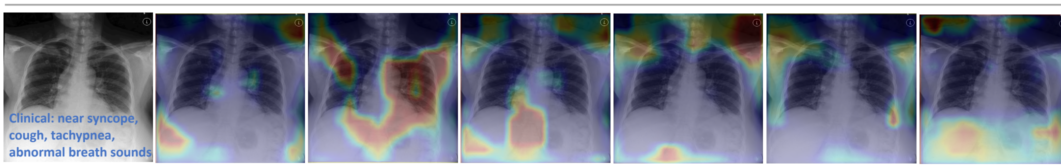
Ours: bilateral small to moderate pleural effusions are presented. minimal patchy opacity in the right lung base is noted possibly an area of pulmonary edema. there is hyperexpansion of the lungs raising the possibility of some underlying chronic pulmonary diseases. linear atelectasis in the left lower lobe. heart size is likely normal. no pneumothorax is present.

Ours-wo-VSMA: there is no evidence of pneumothorax or right pleural effusion. patchy bibasilar airspace opacities may reflect atelectasis. no definite focal consolidation. no acute cardiopulmonary process. cardiomediastinal silhouette is within normal limits. no acute osseous abnormalities.

Ours-wo-DPL: lungs are mildly hyperinflated. bilateral pleural effusions are trace. no focal airspace consolidation or pneumothorax. possible atelectasis is present. the heart size is normal. the hilar and mediastinal contours are normal. there is no pneumothorax.

Ours-wo- $\mathcal{L}_{ISM}$&$\mathcal{L}_{DA}$: no pleural effusion or pneumothorax is seen. no acute intrathoracic abnormality. no focal consolidations concerning for infection are identified. the heart size is stable. a large retrocardiac hiatal hernia is again seen. bony structures are intact. normal chest radiographs.

Ground Truth: frontal and lateral views of the chest were obtained. there is a small right pleural effusion with some fluid seen tracking in the minor fissure and which may be partially loculated. scattered patchy opacities projecting predominantly over the right lung raises concern for an infection less likely asymmetric edema. there is left basilar atelectasis. the lungs are relatively hyperinflated with flattening of the diaphragms suggesting chronic obstructive pulmonary disease. the cardiac and mediastinal silhouettes are relatively stable.



Ours: no pleural effusion or pneumothorax is identified. mild opacity at the left lung base is likely atelectasis. the cardiac mediastinal silhouettes are stable. no focal consolidation is seen. no pulmonary edema is seen. no acute cardiopulmonary process.

Ours-wo-VSMA: the lungs are clear. there is no pleural effusion or pneumothorax. there is no focal lung consolidation. cardiomediastinal silhouette is normal. no convincing sign of pneumonia or edema. osseous and soft tissue structures are unremarkable .

Ours-wo-DPL: heart size and cardiomediastinal silhouettes are normal. mild left basilar atelectasis. there is no pneumothorax. there is no evidence of focal lung consolidation or pulmonary vascular congestion. no edema or pleural effusion.

Ours-wo- $\mathcal{L}_{ISM}$&$\mathcal{L}_{DA}$: there is a collapse with opacification at the left base consistent with pleural fluid and either atelectasis or consolidation. a small pleural effusion has probably resolved on the right side for the most part. there is no evidence of pulmonary vascular congestion. heart size is within normal limits.

Ground Truth: the cardiac mediastinal and hilar contours appear stable. there is no pleural effusion or pneumothorax. there are vague opacities in the lingual and left lower lobe the left lower lobe opacity present before. findings are most suggestive of atelectasis. there is no evidence for pulmonary edema. bony structures are unremarkable. left basilar opacities probably atelectasis.

Figure 1. Illustration of generated sentences by our method and comparisons with baselines. For reports by **Ours**, the key findings correctly mentioned in the report are highlighted by green, and those wrongly described are in **red**. The blue texts are input clinical information.