VividMed: Vision Language Model with Versatile Visual Grounding for Medicine

Lingxiao Luo*, Bingda Tang*, Xuanzhong Chen, Rong Han, and Ting Chen

Tsinghua University

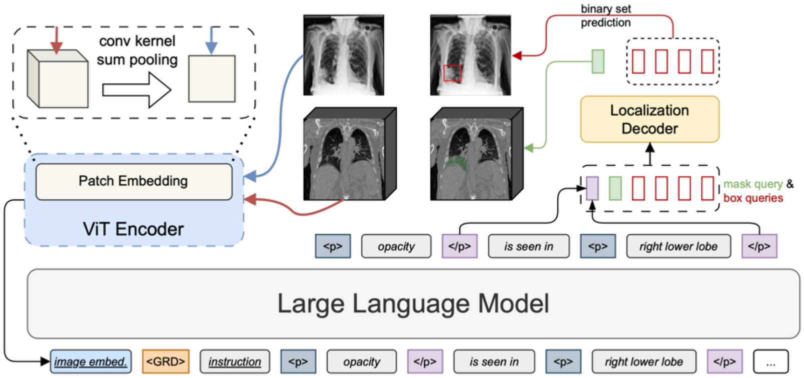
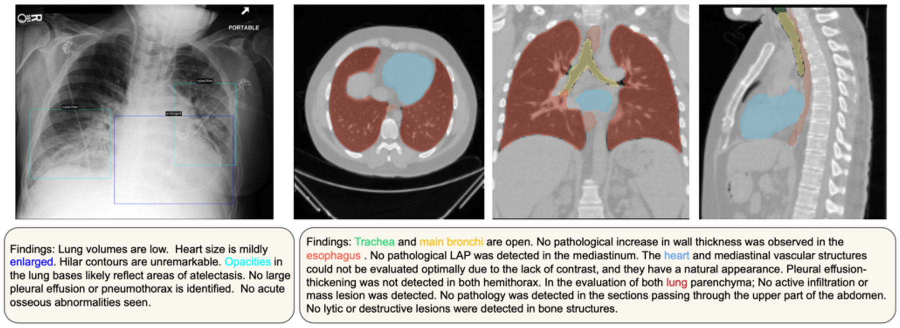


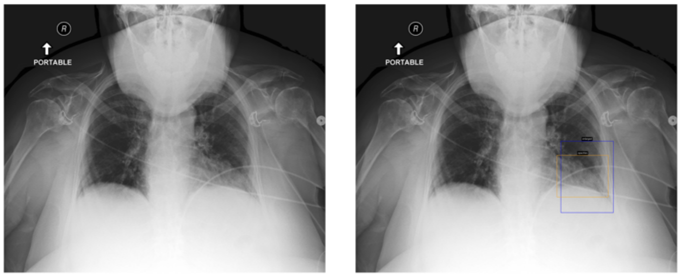
Figure 1: The architecture of VividMed, which is built upon a base VLM (left and lower) and a promptable localization module (upper right). The model identifies key phrases for grounding by enclosing them with bracket tokens, and the hidden states of the closed bracket token is used for prompting the localization module. The model accepts both 2D and 3D images as input by adaptively adjusting weights in the patch embedding layer. The vision encoder of the localization module is omitted for clarity.



Findings: Lung volumes are low. Heart size is mildly enlarged. Hilar contours are unremarkable. Opacities in the lung bases likely reflect areas of atelectasis. No large pleural effusion or pneumothorax is identified. No acute osseous abnormalities seen.

Findings: Trachea and main bronchi are open. No pathological increase in wall thickness was observed in the esophagus . No pathological LAP was detected in the mediastinum. The heart and mediastinal vascular structures could not be evaluated optimally due to the lack of contrast, and they have a natural appearance. Pleural effusion-thickening was not detected in both hemithorax. In the evaluation of both lung parenchyma; No active infiltration or mass lesion was detected. No pathology was detected in the sections passing through the upper part of the abdomen. No lytic or destructive lesions were detected in bone structures.

Figure 2: Selected qualitative results for grounded report generation, zoom in for better view. Impressions are omitted for clarity.



VividMed: Findings: Lung volumes are low. Heart size is mildly enlarged. Mediastinal and hilar contours are unremarkable. Pulmonary vasculature is not engorged. Patchy opacities are noted in the lung bases. No pleural effusion or pneumothorax is present. No acute osseous abnormalities detected.
Impression: Low lung volumes with patchy bibasilar airspace opacities, likely atelectasis in the setting of chronic obstructive pulmonary disease.

Reference: Findings: The lung volumes are low. A patchy left basilar opacity obscures the cardiac border and apex of the left hemidiaphragm, worrisome for pneumonia. Elsewhere, the lungs appear clear. There are no pleural effusions or pneumothorax.
Impression: Left basilar opacity worrisome for pneumonia.

Figure 3: In this example, the model wrongly identifies cardiomegaly and gives an unusual visual grounding result, which may remind the radiologist in clinical practice.