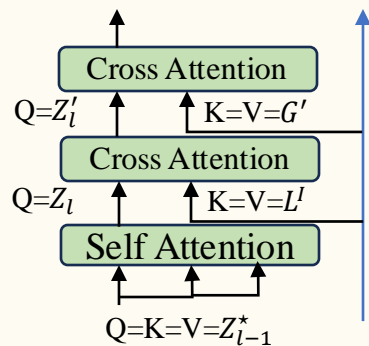
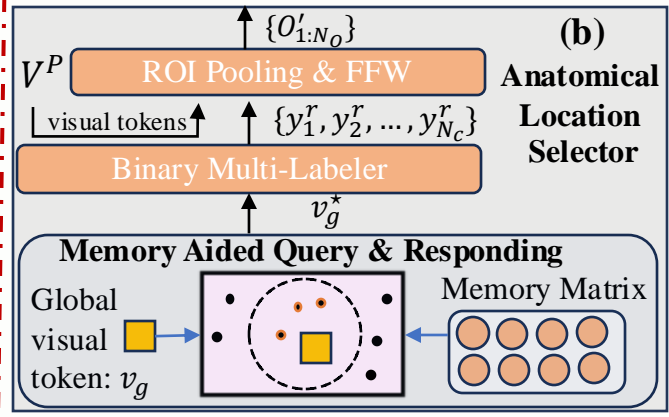


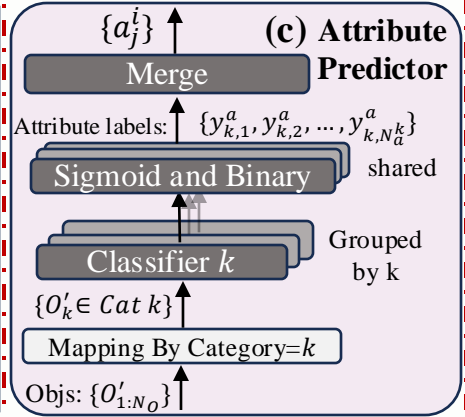
**(a) Scene Graph Aided Decoder**



**(b)**

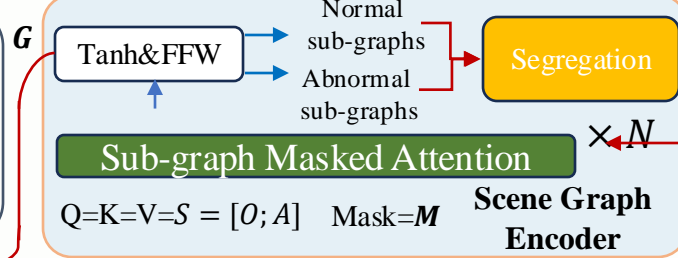
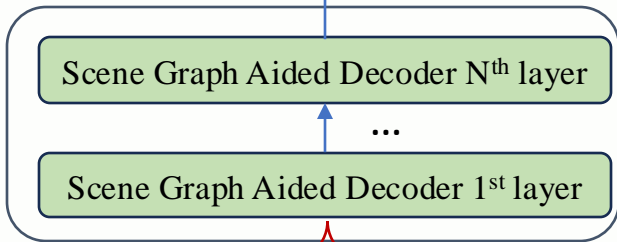


**(c)**



**Scene Graph Construction**

- $S_1$ : left lung
- $S_2$ : right lung
- $S_{N_O}$ : aortic arch
- $S_{N_O+1}$ : lung opacity
- $S_{N_a-1}$ : tortuous aorta
- $S_{N_a}$ : vascular calcification



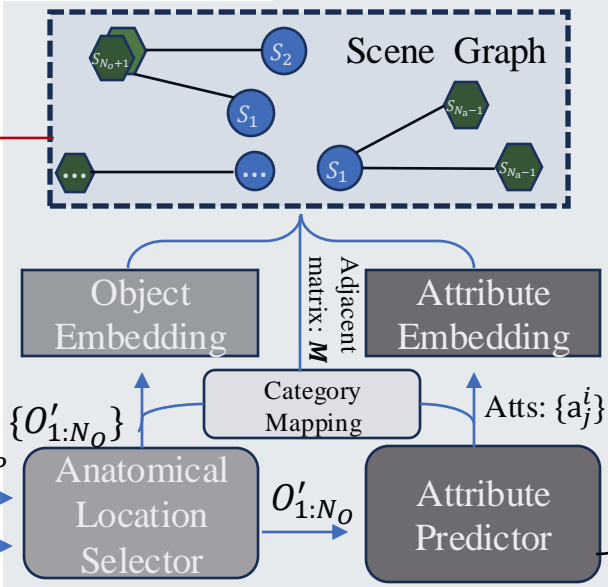
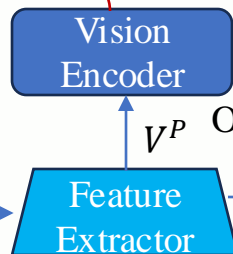
Encoded Textual Tokens:  $Z_0^*$

Report: heart size is normal. cardio-mediastinal silhouette and hilar contours are unremarkable. lungs are clear...

Encoded Visual Tokens:  $L^l$



Images with bounding boxes



**(d) The overall framework of SGRRG**

$G$ : Encoded Scene Graph Embedding