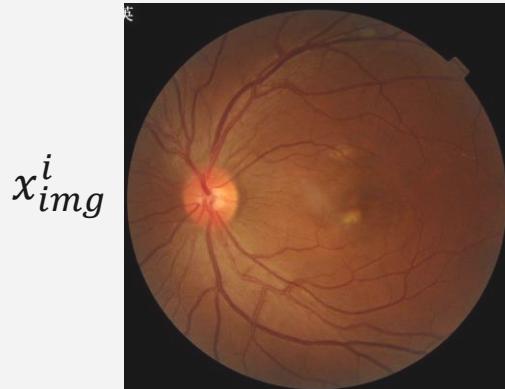


Input Text Question  $x_{txt}^i$



LVLM  
 $\mathcal{M}_{LVLM}$

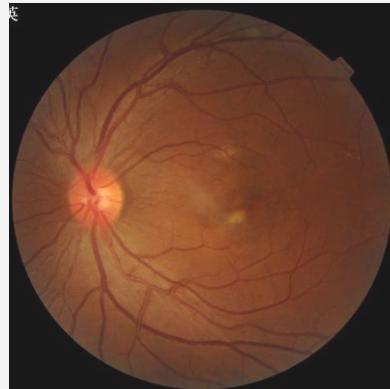


LoRA



It is [SEG]. +diagnosis result

$x_{img}^i$

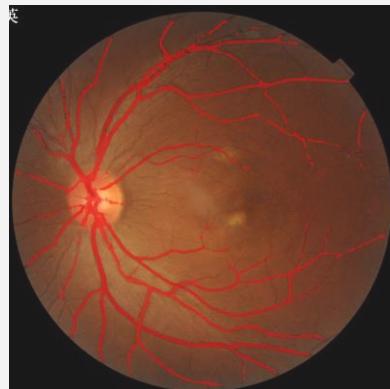


SAM  
Encoder

$\mathcal{M}_{SEG}^{enc}$



$\hat{o}_{mask}$

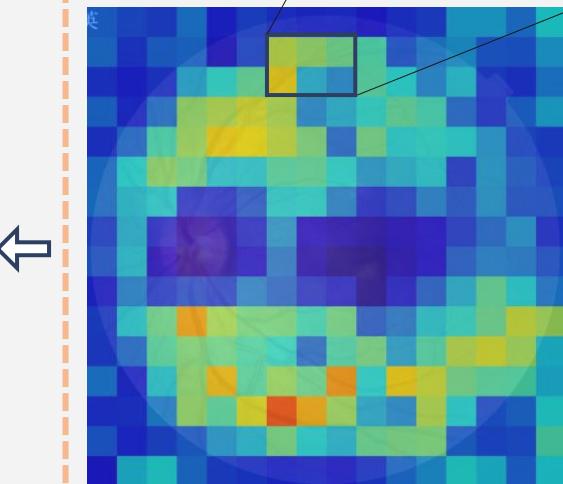


SAM  
Decoder

$\mathcal{M}_{SEG}^{dec}$



Prompt Encoder ( $F, E_{seg}, \mathcal{M}_{region}$ )



$\mathcal{R}$



$\tau$



$\mathcal{M}_{region}$

RVLS2M module

Trainable Frozen