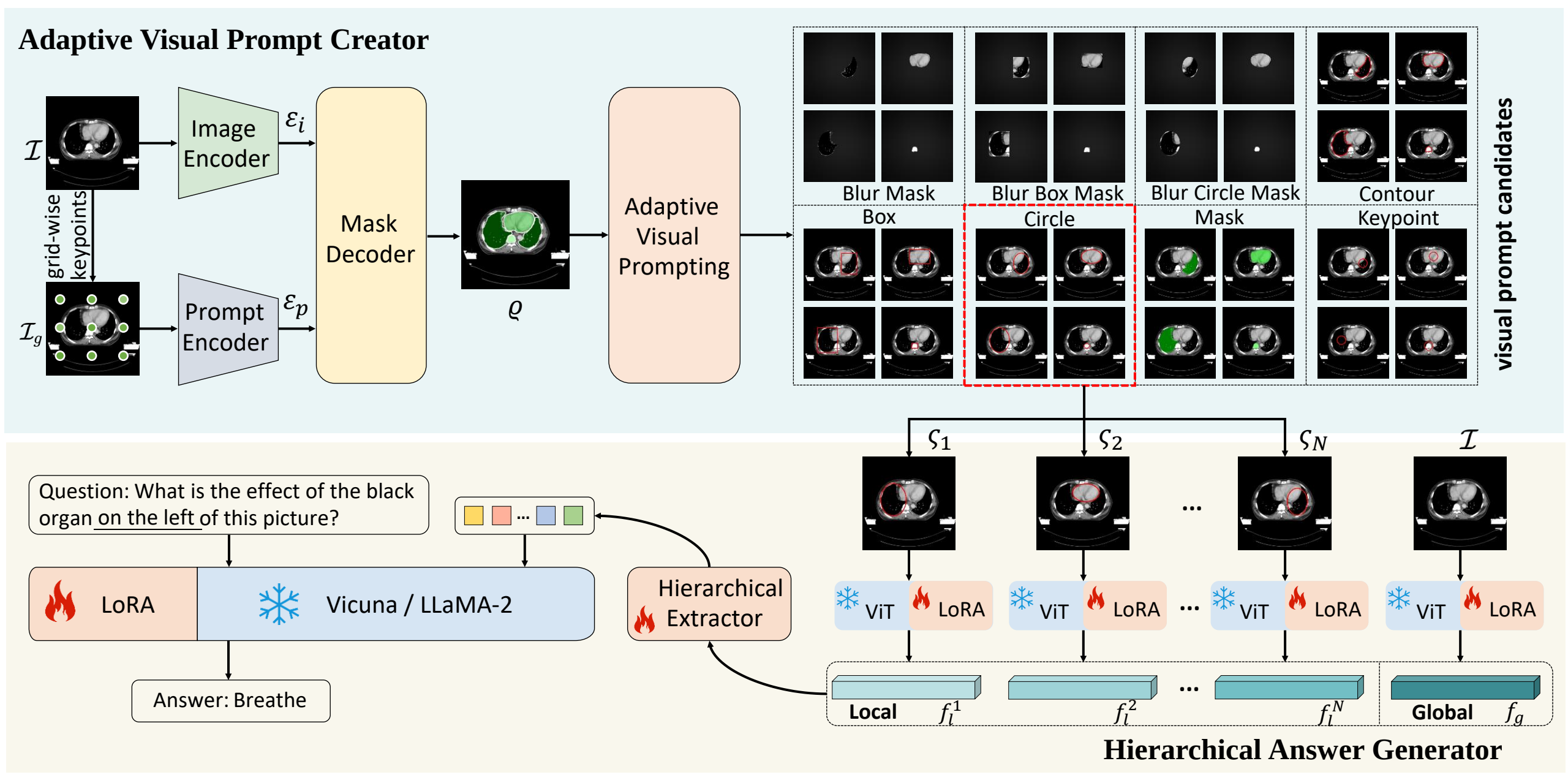
Adaptive Visual Prompt Creator
$\mathcal{I}$
Image Encoder
ε_i
grid-wise keypoints
$\mathcal{I}_g$
Prompt Encoder
ε_p
Mask Decoder
ϱ
Adaptive Visual Prompting
Blur Mask
Blur Box Mask
Blur Circle Mask
Contour
Box
Circle
Mask
Keypoint
visual prompt candidates
ς_1
ς_2
ς_N
$\mathcal{I}$
Question: What is the effect of the black organ on the left of this picture?
LoRA
Vicuna / LLaMA-2
Answer: Breathe
Hierarchical Extractor
ViT
LoRA
Local
f_l^1
f_l^2
f_l^N
Global
f_g
Hierarchical Answer Generator