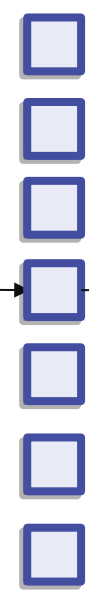
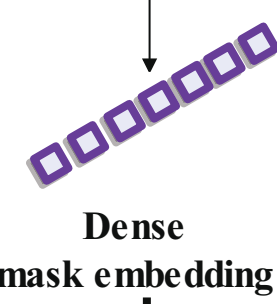
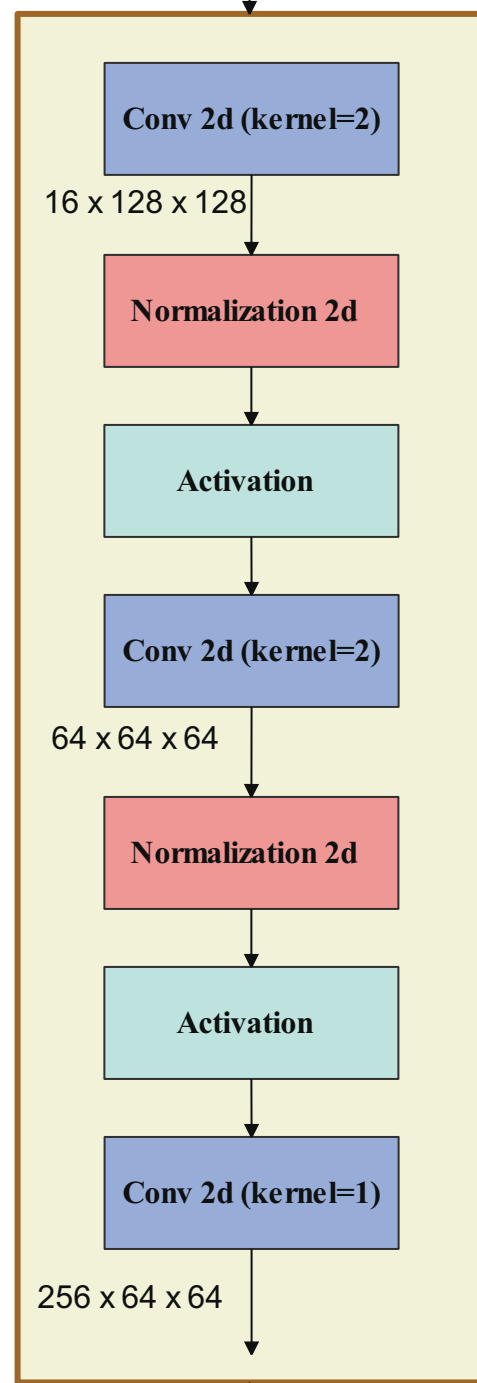


Image embedding

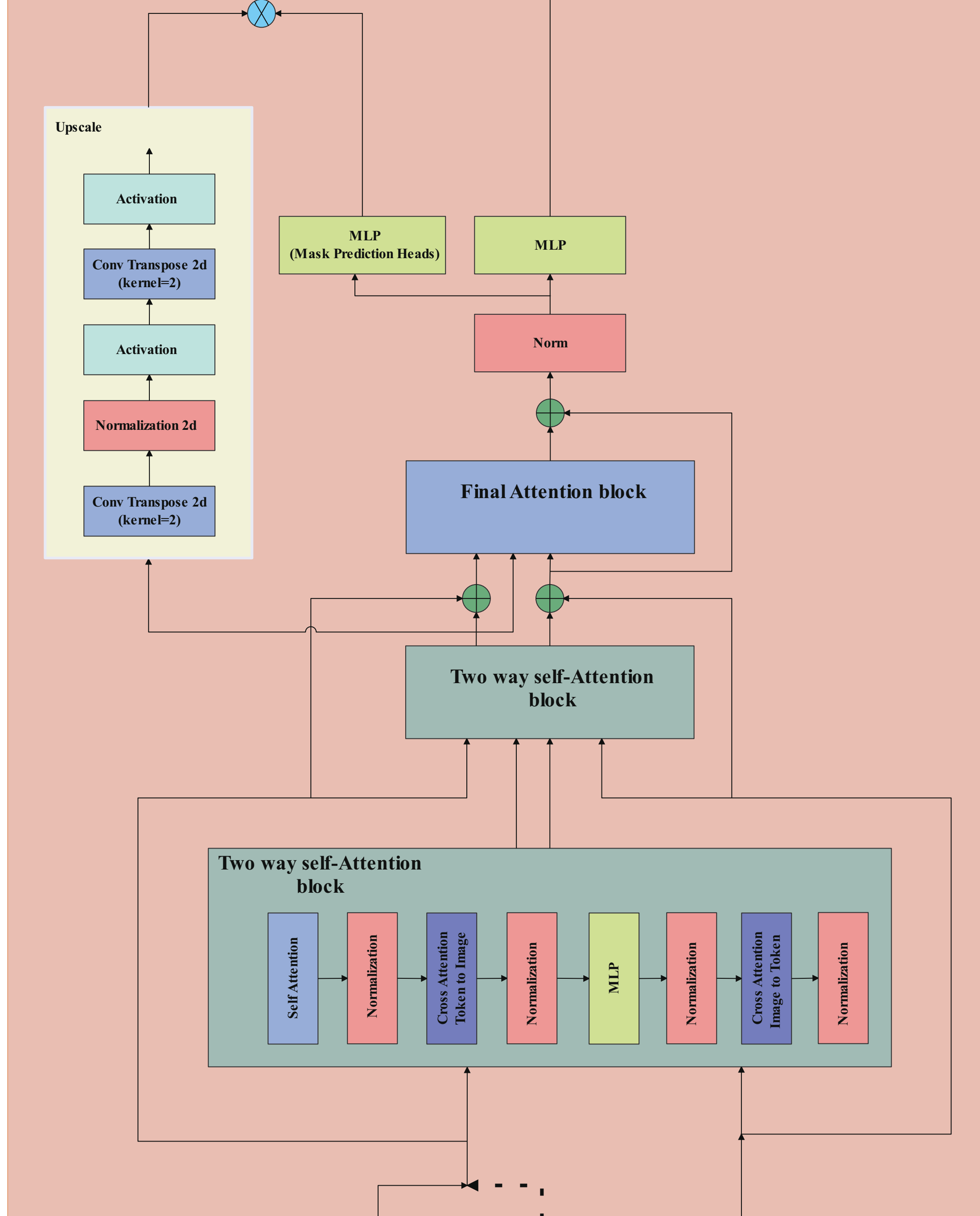


Input mask  
1 x 256 x 256

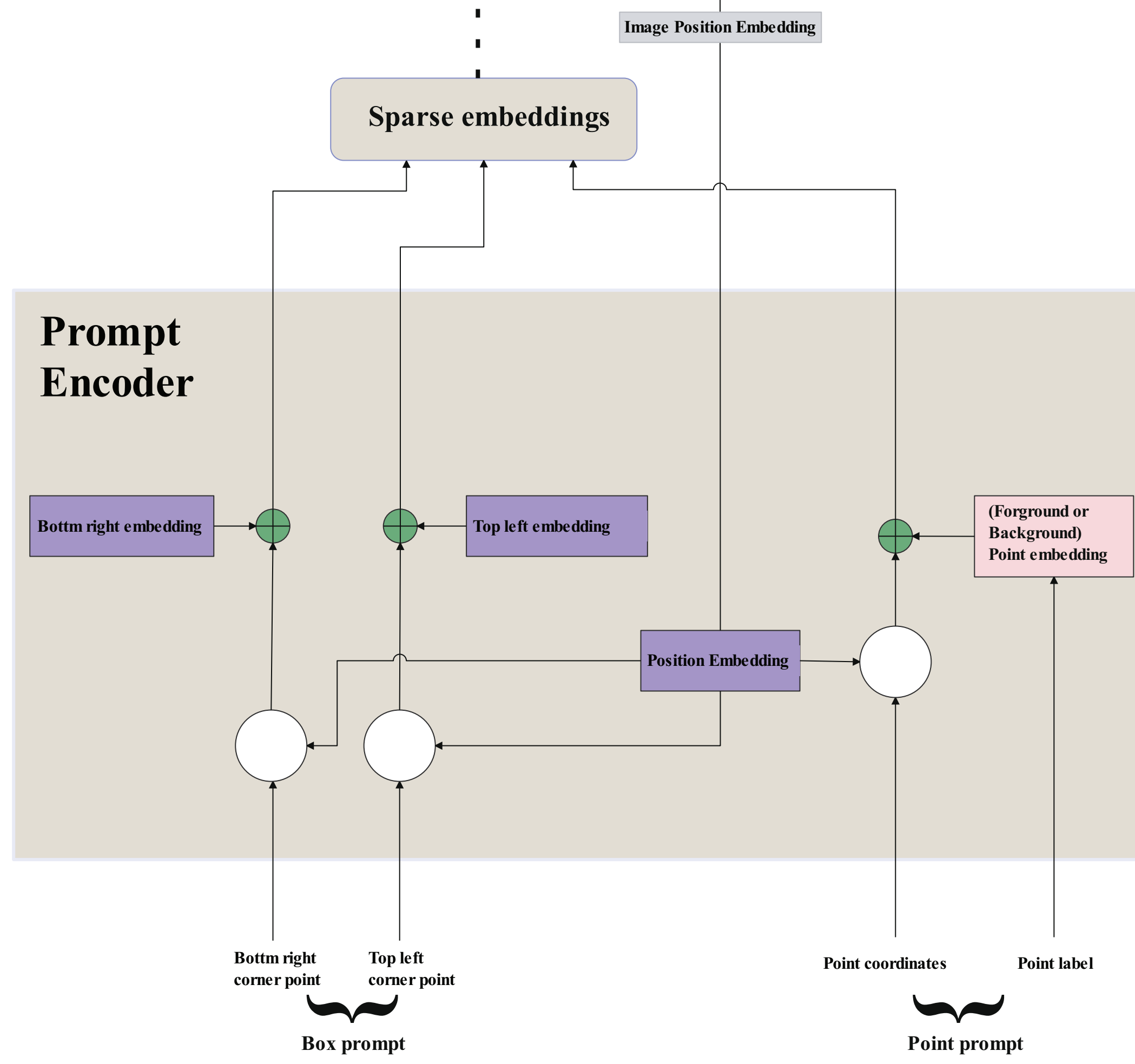


Repeating

Decoder



Prompt Encoder



Preprocess module  
(prompt creator)



Ground truth