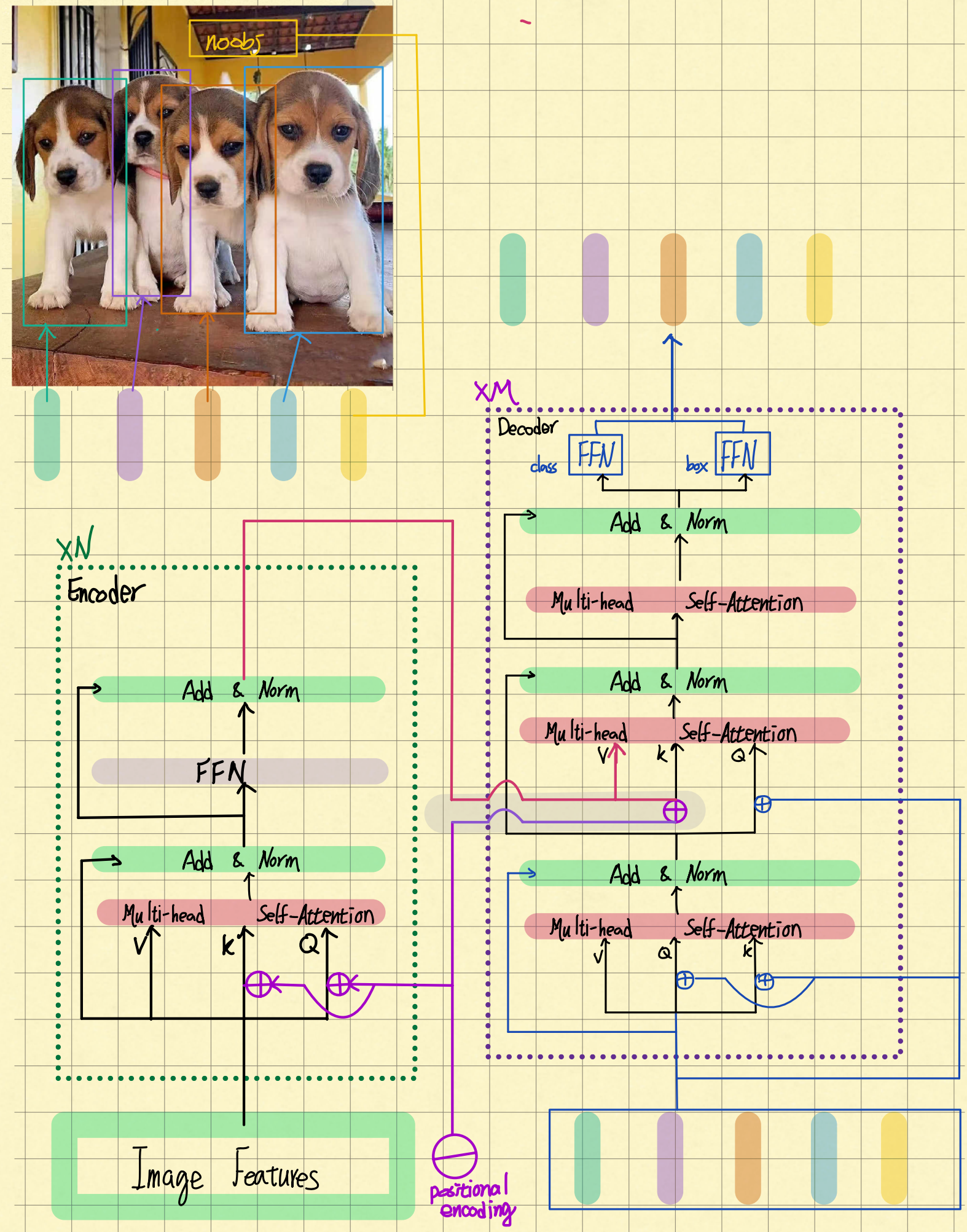
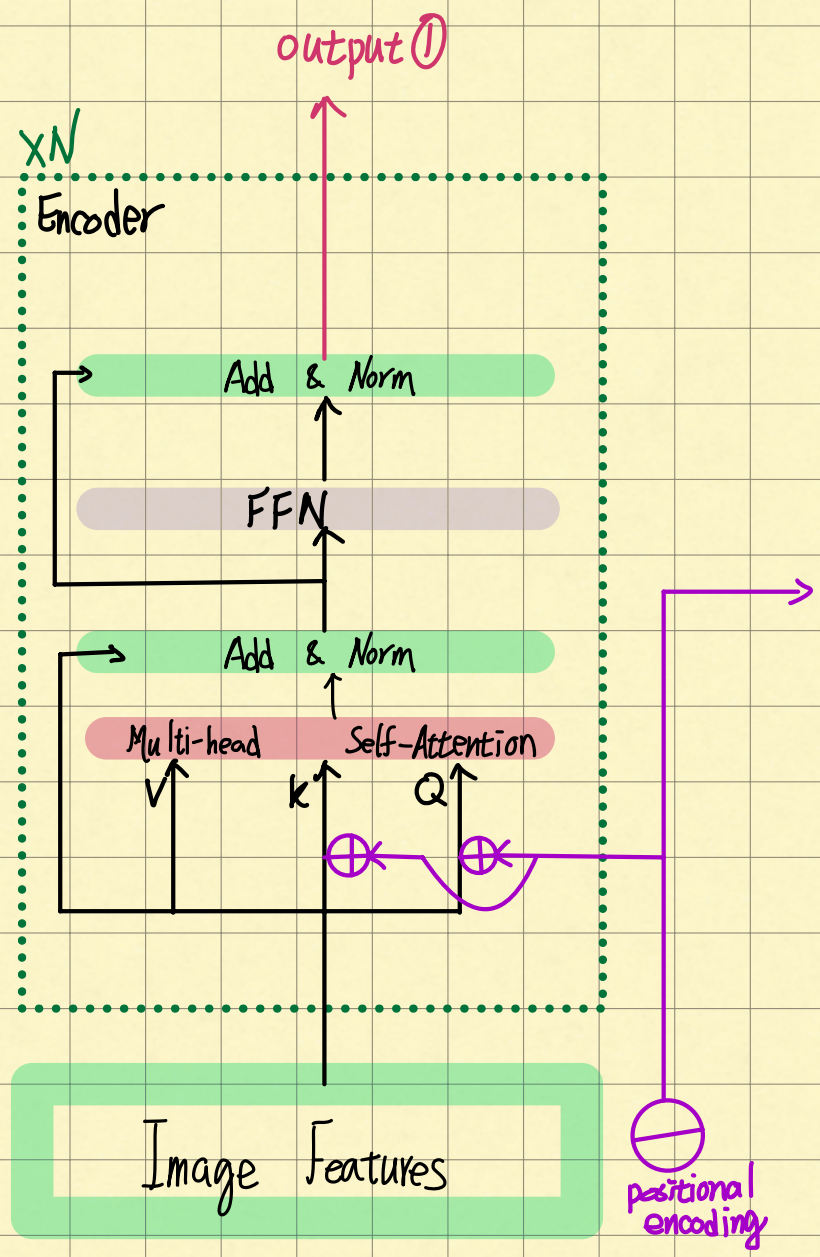

noobj
xM
Decoder
class
FFN
box
FFN
Add & Norm
Multi-head Self-Attention
Add & Norm
Multi-head Self-Attention
V
K
Q
Add & Norm
Multi-head Self-Attention
V
Q
K
xN
Encoder
Add & Norm
FFN
Add & Norm
Multi-head Self-Attention
V
K
Q
Image Features
positional encoding


output ①
xN
Encoder
Add & Norm
FFN
Add & Norm
Multi-head Self-Attention
V
K
Q
Image Features
positional encoding