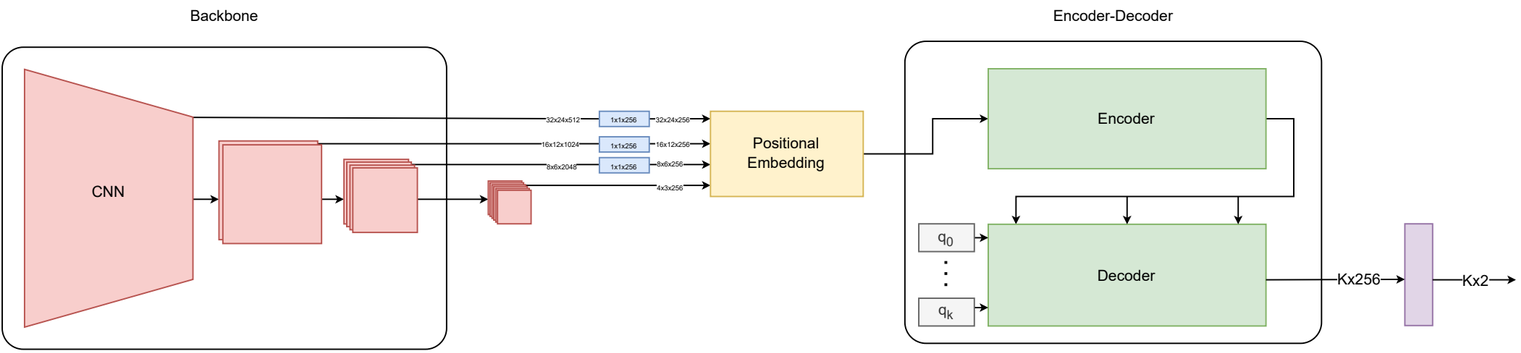
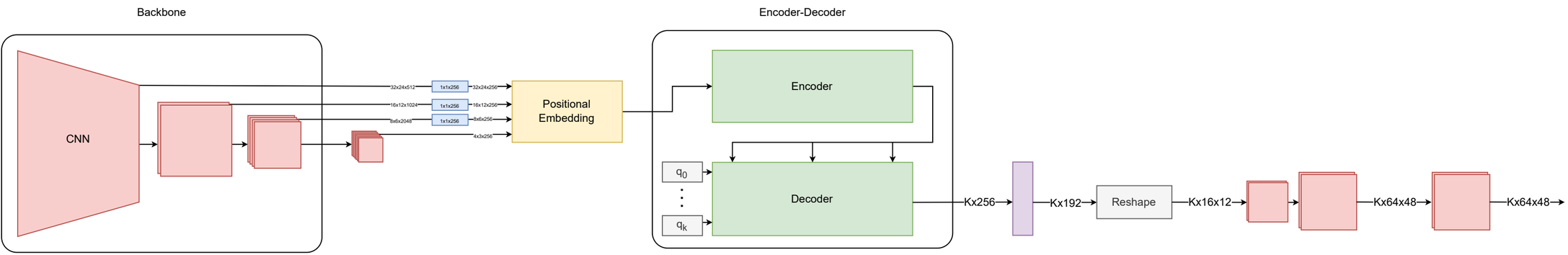
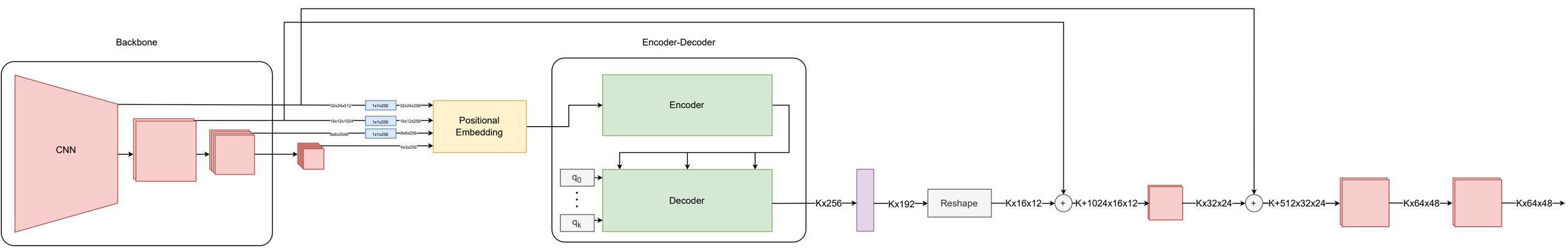

Backbone
Encoder-Decoder
CNN
32x24x512
16x12x1024
8x6x2048
4x3x256
1x1x256
32x24x256
16x12x256
8x6x256
Positional Embedding
Encoder
Decoder
q_0
q_k
Kx256
Kx2


Backbone
Encoder-Decoder
CNN
32x24x512
16x12x1024
8x6x2048
4x3x256
1x1x256
32x24x256
16x12x256
8x6x256
Positional Embedding
Encoder
Decoder
q_0
q_k
Kx256
Kx192
Reshape
Kx16x12
Kx64x48
Kx64x48


Backbone
Encoder-Decoder
CNN
32x24x512
16x12x1024
8x6x2048
4x3x256
1x1x256
32x24x256
16x12x256
8x6x256
Positional Embedding
Encoder
Decoder
q_0
q_k
Kx256
Kx192
Reshape
Kx16x12
+
K+1024x16x12
Kx32x24
+
K+512x32x24
Kx64x48
Kx64x48