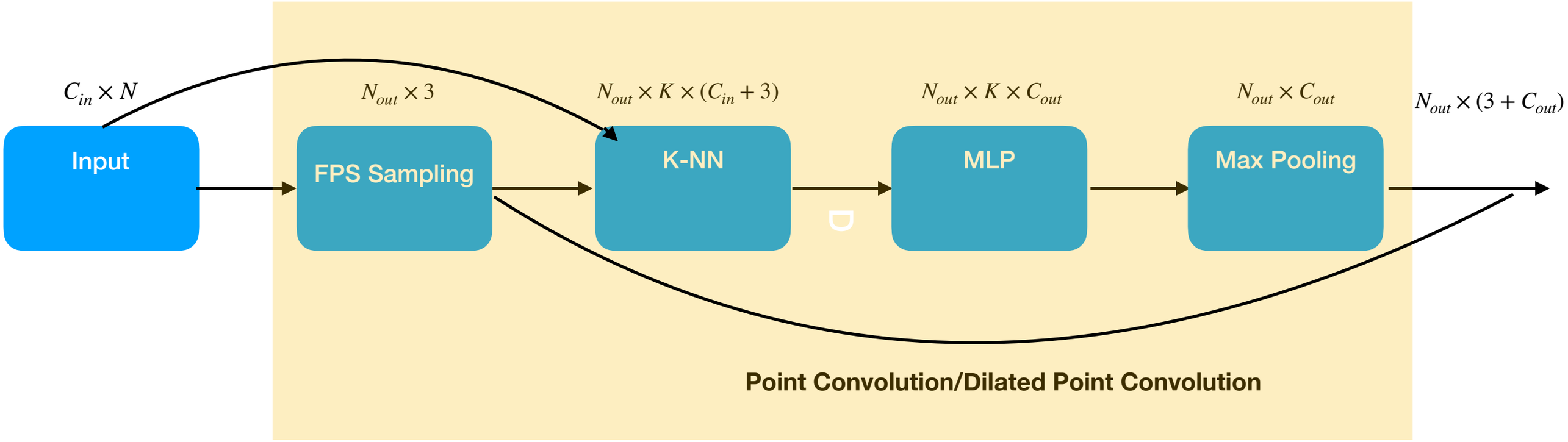
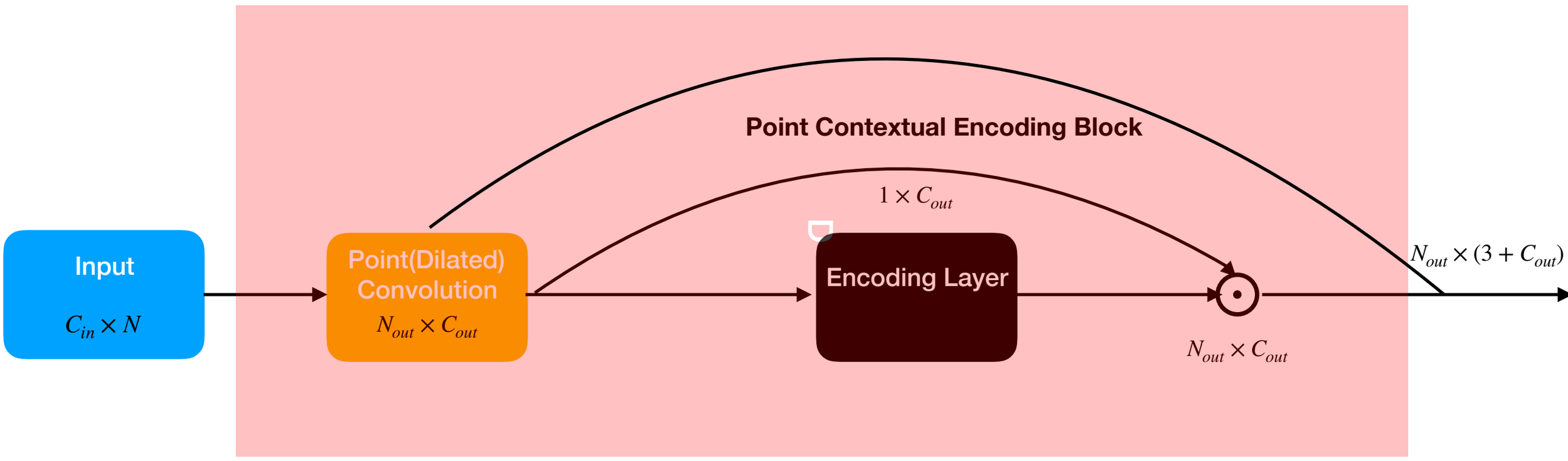

$C_{in} \times N$
Input
$N_{out} \times 3$
FPS Sampling
$N_{out} \times K \times (C_{in} + 3)$
K-NN
$N_{out} \times K \times C_{out}$
MLP
$N_{out} \times C_{out}$
Max Pooling
$N_{out} \times (3 + C_{out})$
Point Convolution/Dilated Point Convolution

(a)


Point Contextual Encoding Block
Input
$C_{in} \times N$
Point(Dilated) Convolution
$N_{out} \times C_{out}$
$1 \times C_{out}$
Encoding Layer
$N_{out} \times C_{out}$
$N_{out} \times (3 + C_{out})$

(b)