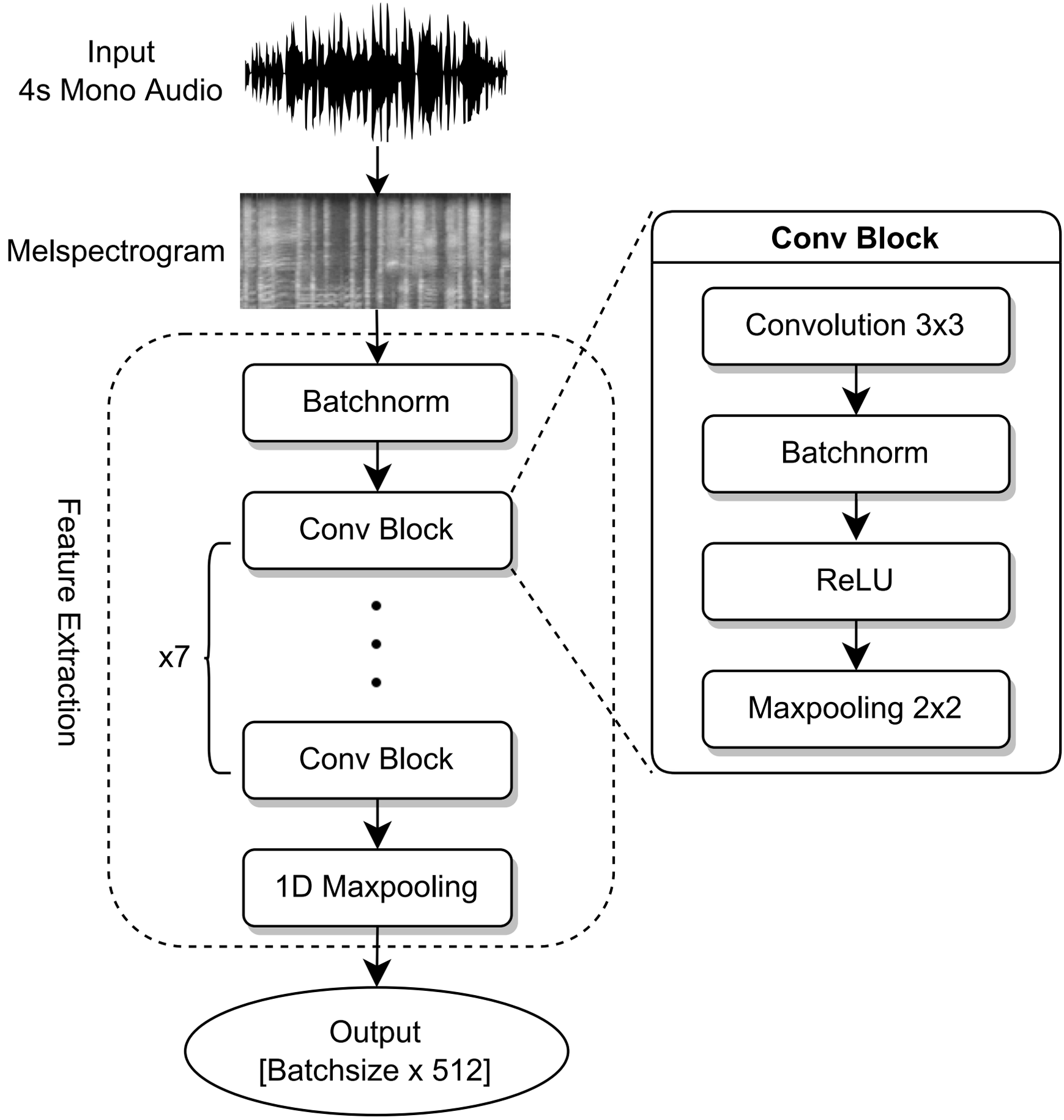
Input
4s Mono Audio
Melspectrogram
Feature Extraction
Batchnorm
Conv Block
x7
Conv Block
1D Maxpooling
Output
[Batchsize x 512]
Conv Block
Convolution 3x3
Batchnorm
ReLU
Maxpooling 2x2