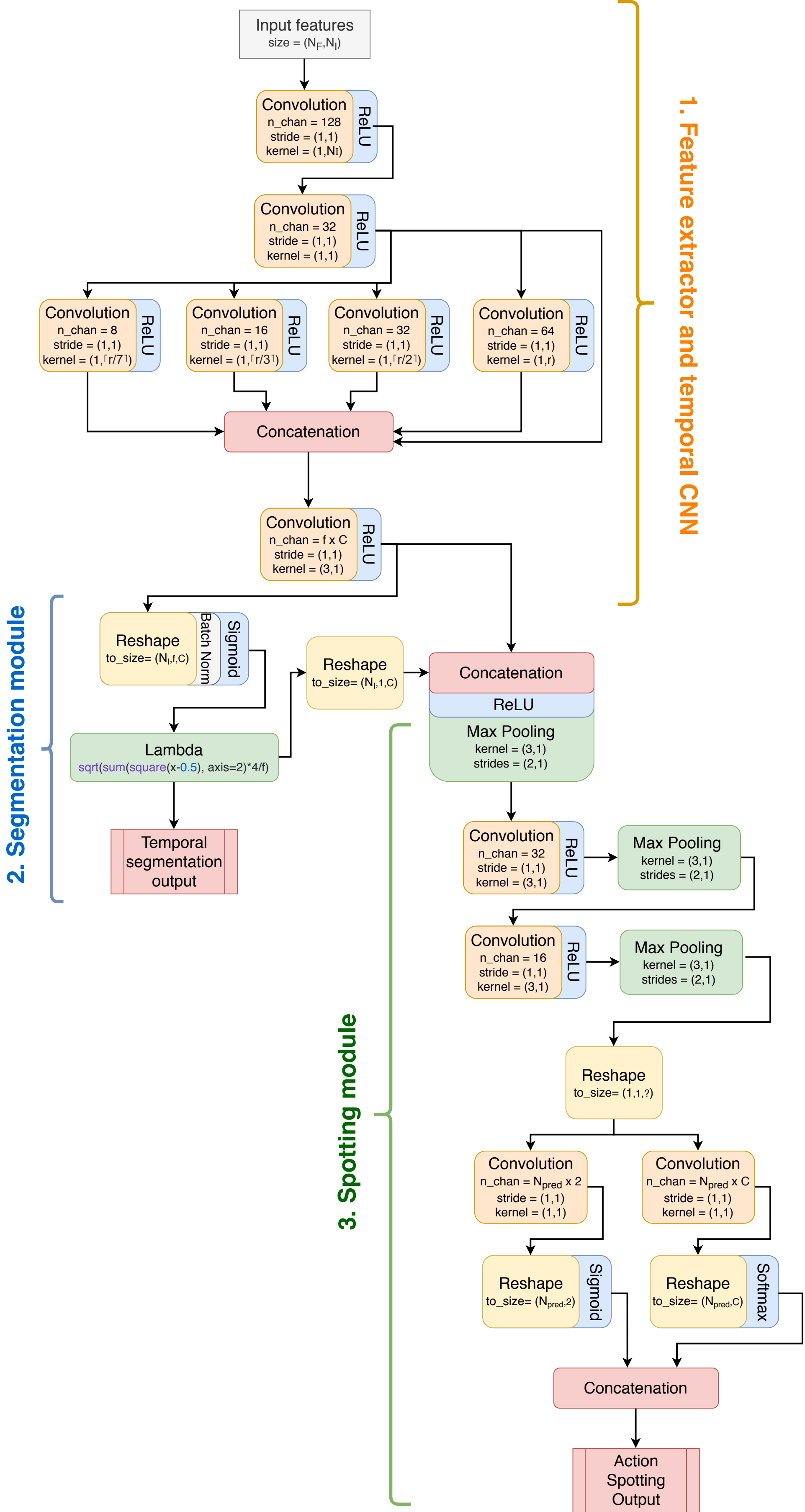
1. Feature extractor and temporal CNN
Input features
size = (N_F,N_I)
Convolution
n_chan = 128
stride = (1,1)
kernel = (1,N_I)
ReLU
Convolution
n_chan = 32
stride = (1,1)
kernel = (1,1)
ReLU
Convolution
n_chan = 8
stride = (1,1)
kernel = (1,⌈r/7⌉)
ReLU
Convolution
n_chan = 16
stride = (1,1)
kernel = (1,⌈r/3⌉)
ReLU
Convolution
n_chan = 32
stride = (1,1)
kernel = (1,⌈r/2⌉)
ReLU
Convolution
n_chan = 64
stride = (1,1)
kernel = (1,r)
ReLU
Concatenation
Convolution
n_chan = f x C
stride = (1,1)
kernel = (3,1)
ReLU
2. Segmentation module
Reshape
to_size= (N_I,f,C)
Batch Norm
Sigmoid
Lambda
sqrt(sum(square(x-0.5), axis=2)*4/f)
Temporal segmentation output
Reshape
to_size= (N_I,1,C)
Concatenation
ReLU
Max Pooling
kernel = (3,1)
strides = (2,1)
3. Spotting module
Convolution
n_chan = 32
stride = (1,1)
kernel = (3,1)
ReLU
Max Pooling
kernel = (3,1)
strides = (2,1)
Convolution
n_chan = 16
stride = (1,1)
kernel = (3,1)
ReLU
Max Pooling
kernel = (3,1)
strides = (2,1)
Reshape
to_size= (1,1,?)
Convolution
n_chan = N_pred x 2
stride = (1,1)
kernel = (1,1)
Convolution
n_chan = N_pred x C
stride = (1,1)
kernel = (1,1)
Reshape
to_size= (N_pred,2)
Sigmoid
Reshape
to_size= (N_pred,C)
Softmax
Concatenation
Action Spotting Output