

A Architecture Details

Operation	Stride	# In	# Out	Activation	Other
Streaming detector					
Featurizer MLP		12	64		With max pooling
Convolution block	1	64	64	ReLU	Layers=4
Convolution block	2	64	128	ReLU	Layers=6
Convolution block	2	128	256	ReLU	Layers=6
LSTM		256	256		Hidden=128
Deconvolution 1	1	64	128	ReLU	
Deconvolution 2	2	128	128	ReLU	
Deconvolution 3	4	256	128	ReLU	
Convolution/Detector	1	384	16		Kernel= 3×3
Convolution block					
(S, C_{in}, C_{out}, L)					
Convolution 1	S	C_{in}	C_{out}	ReLU	Kernel= 3×3
Convolution 2, ..., $L - 1$	1	C_{out}	C_{out}	ReLU	Kernel= 3×3
Normalization Batch normalization before ReLU for every convolution and deconvolution layer					
Optimizer Adam [29] ($\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$)					
Parameter updates 40,000					
Batch size 64					
Weight initialization Xavier-Glorot[18]					

Table 3: Network architecture and training hyper-parameters.