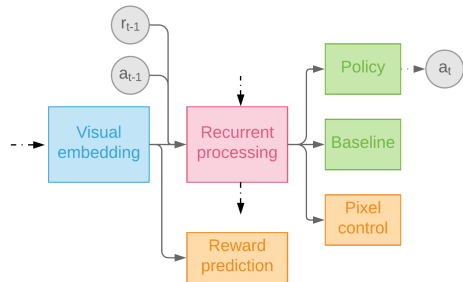
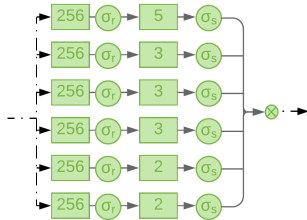


(a) Agent



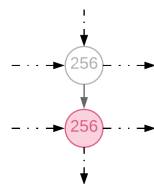
(b) Policy



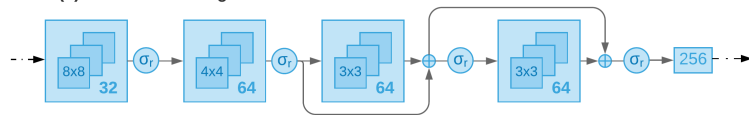
(c) Baseline



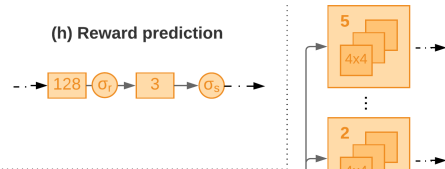
(e) Recurrent processing with LSTM



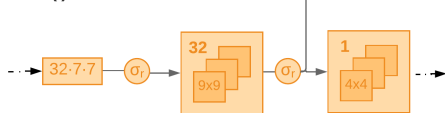
(d) Visual embedding



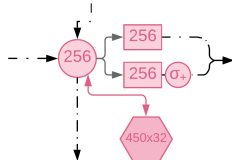
(h) Reward prediction



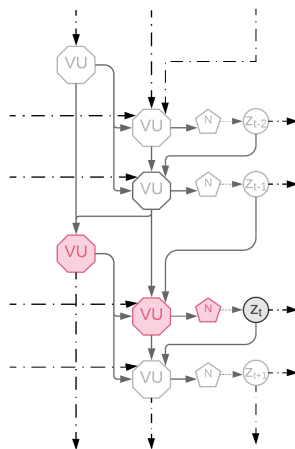
(i) Pixel control



(g) Variational Unit (VU)



(f) Recurrent processing with temporal hierarchy



Legend



Convolution with X KxK filters



Deconvolution with X KxK filters



LSTM with X neurons



DNC memory (LRU) with N slots of size K



Diagonal Normal distribution



Linear layer with X neurons



Linear layer with X.Y.Z neurons reshaped to 3D tensor



ReLU non-linearity



Softmax non-linearity



Softplus non-linearity



Pointwise addition



Outer product

Module
Input/Output

Sampling

Concatenation