**LSTM + CNN:** trained with Adam (learning rate = 0.001) and L2 regularization on all the weights except for the bias with gamma = 5 \* 10^-6

Output size: 200

L2 normalization |

visual vector

