Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering: trained with Adam (learning rate = 0.001) and L2 regularization on all the weights except for the bias with gamma = 5 * 10^-6