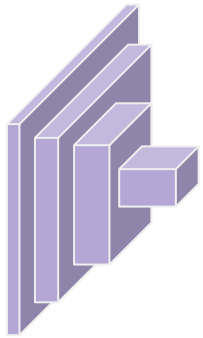
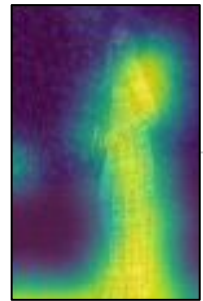




Input Image



StyleNet
(First six convolution layers)



Feature Map(f_l)

Localization Network, ST-LSTM focuses on different clothing items present in the image at each time step.

