attention loss on AIFB dataset during training epochs (bisim)