Attention Block - 1.0 - 0.5 l o s Logit Difference - -1.0 - -1.5 Layer