contrastive loss similarity i_1, i_2, i_3, \dots t_1, t_2, t_3, \dots **Image** text encoder encoder T_1, T_2, T_3, \dots I_1, I_2, I_3, \dots