对比损失

clip　　softmax 得到 概率

$$\mathcal{L}_{t2i}(i) = -\log \frac{\exp(s(V_i, T_i))}{\sum_{a=1}^{B} \exp(s(V_a, T_i))}$$

文本对图像的对比损失

即一个确定的文本对所有图像的对比损失

对 clip 来说，找到(图像-文本)对　(positive)
匹配的

得 $\exp(s(v_i, T_i))$

此文本对所有图像求相似度得 $\sum_{a=1}^{B} \exp(s(v_a, T_i))$

$\mathcal{L}_{t2i}(i)$

$= \log \sum_{a=1}^{B} \exp(s(v_a, T_i)) - s(v_i, T_i)$

越小越好　　越大越好

clip-reid　stage 1.

$$\mathcal{L}_{t2i}(y_i) = \frac{-1}{|P(y_i)|} \sum_{p \in P(y_i)} \log \frac{\exp(s(V_p, T_{y_i}))}{\sum_{a=1}^{B} \exp(s(V_a, T_{y_i}))} \quad (4$$

理解一个文本可对应多张图片.(多个 positive 对)

则对多个 positive 对求平均　　eg (v_1, T_1)

(v_2, T_1)

(v_{10}, T_1)

图片 1. 2. 10 都与 T1 对应

stage 2.

$$\mathcal{L}_{i2tce}(i) = \sum_{k=1}^{N} -q_k \log \frac{\exp(s(V_i, T_{y_k}))}{\sum_{y_a=1}^{N} \exp(s(V_i, T_{y_a}))}$$

图像对文本的交叉熵

图像时各类、文本的得分

$$\text{记为} \quad q_1 \log \frac{\exp(s(v_i, T_{y_1}))}{\exp(s(v_i, T))} + q_2 \frac{\exp(s(v_i, T_{y_2}))}{\exp(s(v_i, T))}$$

q_1 → [1 0 0 0 0 ···] 标签平滑

q_2 → [0 1 0 0 0 ···]

$\frac{\exp(s(v_i, T_{y_1}))}{\exp(s(v_i, T))}$ → 第一张图片 第一条文本

其中 q_k 代表的图像 与 T_{yk} 代表的文本是成对的

v_i

使 $v_i = T_{y_i}$ 时最大

q_1 [1, 0, 0, 0 ···]

$\frac{\exp(s(v_i, T_{y_1}))}{\exp(s(v_i, T))}$ → [0.9 0.01 0.01 ···]

直观解释：

对一张图片 v_i 把它与每个类别文本特征的相似度做 softmax 得到预测概率分布，再用 q_1（真实、平滑标签分布）作交叉熵。