**NFormer:使用Neighbor Transformer行人重识别具有鲁棒性**

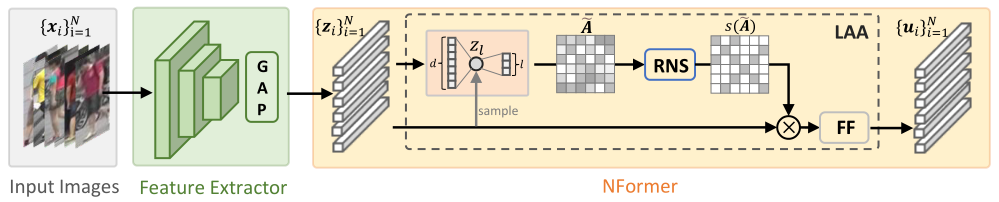
**摘要：**

人物再识别旨在通过不同的摄像机和场景在高度不同的环境中检索人物，在这种情况下，健壮且有区别的表征学习至关重要。大多数研究都考虑从单个图像中学习表征，忽略了它们之间的任何潜在交互作用。由于建模大量图像之间的交互是一项具有大量干扰因素的艰巨任务，NFormer引入了两个新模块，即Landmark Agent Attention和Reciprocal Neighbor Softmax。

**贡献：**

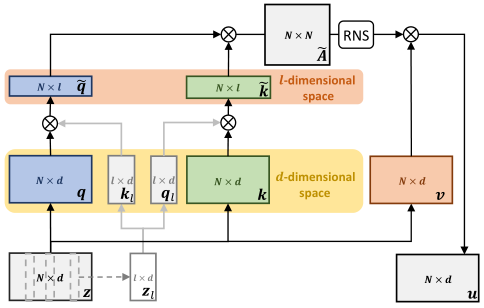
1. 提出了一个 Neighbor Transformer Network (NFormer)，以有效地在训练和测试时建模所有输入图像之间的关系
2. 提出了一个 Landmark Agent Attention (LAA)，通过在表示空间中引入少量landmark agent来减少亲和矩阵的计算量
3. 提出Reciprocal Neighbor Softmax (RNS)函数，来实现稀疏attention，只关注计算上可管理的邻居。RNS显著地约束了不相关个体之间的噪声交互，使表示聚合过程更加有效和高效

**主体架构：**

GAP: 全局平均池，LAA: Landmark Agent Attention， RNS: Reciprocal Neighbor Softmax，FF: 前馈网络

整个架构主要由两部分组成：特征提取器和NFormer。其中，NFormer主要由两部分组成：Landmark Agent Attention (LAA)和Reciprocal Neighbor Softmax (RNS)。

**Landmark Agent Attention**



如果依照传统的Attention进行计算，由输入Z得到,这样在计算 q 和 k 获取相似性矩阵时复杂度较高，为O(N2d)。 为此，作者如下改进：(用m代替图中的l方便区分数字1，下面公式中m和l一致)

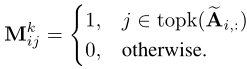
1. 在输入 z 中随机采样m个样本得到zm，然后生成km和qm，这样特征就从N×d降为m×d。
2. 将原始的 q 和 k 通过与km和qm分别相乘，得到和。
3. 和计算得到 N×N 的相似性矩阵 。通过上述操作，将复杂度就从O(N2d)降低为O(N2m)。根据文中描述，m=5,d=256，显著的降低了计算量。



**Reciprocal neighbor softmax (RNS):**

原始的Softmax计算是聚合所有的样本，但是不相关样本的显著存在会对最终计算产生负面影响。同时，Softmax的计算复杂度也是 O(N2d) 的，因此作者认为对于每张图片只需要取与他最具有相似性的前k张图片进行特征聚合即可。

作者提出Reciprocal Neighbor Softmax (RNS)，用reciprocal邻居掩码强制对少数相关的attention weight进行稀疏化。根据这一操作，作者首先根据  将每行的前k个最大的位置设置为1，其他位置设置为0，得到 Mk。



将Mk与其转置相乘，就可以得到一个掩码矩阵M：

