|  |  |  |
| --- | --- | --- |
| 文章信息 | 摘要 | 方法 |
| **阅读程度**：  精读(跟视频)  **文章标题：**  Momentum Contrast for Unsupervised Visual Representation Learning  **中文标题：**  使用动量对比进行无监督视觉表征学习  **发表于：**  CVPR2020  **作者：**  Kaiming He  **单位：**  Facebook AI Research (FAIR) | 我们提出了动量对比（MoCo）用于无监督的视觉表征学习。将对比学习视为字典查找问题，从这个角度出发，我们使用队列与移动平均编码器构建了一个动态字典。这使得一个大的、一致性的字典得以被构建，从而促进无监督学习。  MoCo在ImageNet分类的通用线性协议下提供了有竞争的结果。  更重要的是，MoCo学习到的表示可以很好地转移到下游任务中。MoCo可以在帕斯卡VOC、COCO和其他数据集上的7个检测/分割任务中优于其有监督的训练前对手，有时会大大超过它。这表明，在许多视觉任务中，无监督和监督表征学习之间的差距在很大程度上被缩小。 | 第一个视角是介绍Moco这个操作：  首先，进行相同的初始化  抽取一个minibatch即，进行两种随机数据增强得到，，分别通过，得到，。  ，均为形状为的矩阵，第一维是minibatch样本数，第二维是特征向量的维数，第二维对位相乘得到，形状为，即N个正样本对特征之间的余弦相似度。  的形状为，是从队列中取得K个负样本，队列遵循先进先出原则，即一个新的minibatch算得后，将这一批送入队列，将最旧的那一批移除队列，保持队列的长度不变，本质上队列也类似于一个memoreybank，区别在于memoreybank的负样本是从整个训练集中抽样，会选到很久以前取得的特征向量，而使用这种先进先出的队列保证了负样本来自最近处理过的几批minibatch，这样就保证了更好的**一致性（即负样本由同一个或者差别很小的****提取到的特征向量）。**  之后使用infoNCE求得损失函数，使用反向传递更新，而使用动量更新，每次为0.999的上一代加上0.001的最新版。  说实话，这样缓慢更新的确实生成了质量不错的自监督信号，属实是敢于操作了。    关于三种方法透彻的总结  第一种：我看过的方法中Invariant Spreading是一个很好的例子，即为了保证**一致性，**所有负样本均来自当前minibach，自然字典是当前minibatch基于当前encoder k获得的，这保证了绝对的一致性，但是minibatch的大小受硬件的限制，不能太大，这限制了负样本数量的**大**小。  第二种：我看过的方法中Inst Dis及CMC都是这样基于memorybank的一种形式，这种形式负样本的数量受硬件的限制很小，memorybank中的特征向量数就等于训练样本数，但是memorybank中每次更新的特征向量数等于batch大小，如果一个epoch有1000个batch，memorybank中的特征向量有来自1000次更新之前的特征提取网络，有的来自最新的特征提取网络，这显然限制了特征向量（或者说字典的一致性）。  第三种：就是Moco这种，使用先进先出FIFO这种形式的queue（队列），队列中只含有最新的几个batch，同时使用更新缓慢的动量编码器，很好地保证了queue中特征向量的**一致性，**同时queue的大小跟minibatch的大小彼此不再相互限制，从而负样本的数量不再受硬件的影响。 |
| 介绍 |
| **第一段**：介绍了自然语言领域相比与计算机视觉领域需要对图片进行字典的构建，这显然比对单词进行字典构建要复杂的多。（大环境描述）  **第二段**：介绍已有的视觉领域的方法使用对比损失，源于各种动机进行无监督学习，归根结底就是图片信息的字典查询问题。（已有方法总结）  **第三段**：总结了字典构建的两个关键：大，一致性。（指出待解决或待优化的问题）  **第四段**：简述Moco用队列来实现大，用动量编码器来实现一致性。（介绍提出的方法怎么解决）  **第五段**：使用InstanceDIS代理任务在分类问题中成果显著（初步展示）  **第六段**：在其他七个下游任务中也很厉害（进一步展示更多优势） |