利用卷积神经网络学习和迁移中等级的图像表示

首先作者肯定了卷积神经网络在图像分类中的出色表现，并且将其归因于中等层次的图像表示，但其需要调整大量的参数是卷积神经网络的受训练数据限制的原因。（对所研究领域的肯定与总结，并指出其限制，引出本文的工作）。

这里作者主要展示了如何将大规模注释数据集上的CNN学习到的图像表示迁移到有限规模的训练数据中，作者在这里涉及了如何将ImageNet上训练图层迁移到PASCAl

VOC数据集中去，并展示了这样做的效果。（提出作者的主要工作和成果）。

作者认为过去几年图像识别任务的进展主要是由于特征设计水平的提高（HOG和SIFT），以及高质量数据集的提出（Caltech256，Pascal VOC，ImageNet）。

这里作者介绍了一下神经网络的发展。

再次强调卷积神经网络的参数多，应用广。

值得注意的是许多成功的图像分类方法分享了卷积神经网络和神经认知机的概念，一些利用局部特征聚合和神经认知机设计出的低级图像特征效果可能超过CNN，CNN很难用少量特征进行设计。

大数据集提高了图像分类的准确率，但问题是是否解决每一个问题都需要大数据集。有人认为数据集在图像统计上存在差异，如Caltech256和ImageNet通常以目标为中心，其他数据集如Pascal VOC和Label Me通常是背景中包含目标，视角的差异，背景等因素的影响会影响到跨领域的训练和测试时的识别性能，其他领域也有类似的差异，如NLP，困难就是数据量有限

作者提出了将大数据集下使用CNN学习到的图像表示转移到其他视觉识别任务中，特别地，作者设计了经过ImageNet训练的CNN层来获Pascal VOC中图像的中间表示。

这里涉及迁移学习，视觉目标检测，深度学习。迁移学习目的时在源域与目标域间传递只是。在机器视觉中，迁移学习利用调整其它类别训练的分类器来克服某些类别训练数据不足的问题，还有的方法处理相同类别的源域和目标域中不同的数据分布，例如由于照明、背景和视角的变化，通常通过标准图像特征来调整分类器。与这些不同的是作者是要将原任务中的图像表示进行迁移。

与我们工作更类似的是在无监督的任务上进行CNN训练作者解决了在大规模数据上预训练卷积层的任务，并且解决了在图像的尺寸和目标的变化的问题，NLP问题的解决也与卷积神经网络的迁移学习有关系。

作者认为物体识别的关键是一些中间层次的图像表示，卷积神经网络在可以提供这方面的迁移学习。

CNN特征的迁移

这里工作的关键思想是将CNN的内部层作为图像表示的通用提取器，可以在一个数据集上进行预先训练然后使用与目标任务。然而源数据集和目标书记图像的标签和分布十分不同时很困难，本文将对这一问题进行解决（将原任务和目标任务的标签进行了重新映射）和滑动窗口检测器（Faster RCNN表现更好）。

作者应用了滑动窗口的策略，在每个图片上提取了500个正方形的patches，每个patch的宽度为s = min(w, h)/拉姆达，w,h是图像的宽和高，拉姆达属于{1,1.3,1.6,2,2.4,2.8,3.2,3.6,4}，然后重归一化为224\*224的图像大小。

采样的图像可能包含一个或多个背景或是只有目标的一部分，为了标注训练图像patch，作者测量了box与ground truth的重叠程度，这里将这方面分成不同的方面进行讨论，重叠程度大于0.1的，重叠程度大于0.6，重叠了不只一个物体的。这里有一个额外的背景标签，这里还有一个问题是数据不均衡，这里通过随机采样训练patch来实现。

实验

作者这里首先在ImageNet进行预训练，然后在VOC2007和VOC2009上进行迁移训练，最后在一个新的任务上进行目标迁移学习。

1.训练卷积神经网络

周期性的在训练集和验证集上进行训练，初始学习率是0.01,在训练过程中学习率除以10，没有为这个任务预训练参数，未观察到过拟合。

2.ImageNet训练一周达到预期学习率

3.在VOC2007上有个明显的的提高

4.在VOC2012上有个明显的的提高

这里作者探究了变化源域和目标域目标类别重合程度对实验效果的影响，显示会有轻微的影响，这可能是由于训练数据减少影响的，这证明了作者的这一方法具有较好的稳定度。

作者增加了ImageNet与Pascal VOC特定类的重叠，增大了连接层的尺寸，通过这一改进作者在VOC2012上取得了更好的实验效果，特别是与VOC相关的类别取得了进一步的改进，这表明增加源域图像和类的数量可能会产生更积极的效果。

作者通过实验发现作者的这种方法在目标检测领域似乎也有所应用。（事实也证明了这一点）

作者在动作检测领域上也进行了实验，这里目标框已知，标注其类别，作者拷贝了Iamge Net预训练模型的C1到FC6，同时增加了FCa和FCb层，增加了其表现。

错误分析

作者发现这里的错误时由于一些与目标对象非常相似的样本造成的，并认为可能需要更高级的场景解释。也可能无法识别空间上共生的对象（椅子上的人），可以通过改变训练目标来实现，如允许每个目标有多个标签，这里由于稀疏采样造成非常大和非常小的对象也可能识别。

总结

作者提出了一个迁移学习的方法，使目标任务在源域任务上得到增强。我们计划公开提供中级代表，以期促进CNN与计算机视觉领域发明的复杂技术的结合，如有效索引[19]或可变形模型[14]