Convolutional-RecursiveDeepLearning for3DObjectClassiﬁcation

卷积递归深入学习三维对象分类

摘要：3D扫描技术的最新进展使得可以简单地记录色彩和深度图像，这些图像一起可以改善物体识别。 目前大多数方法依赖于这个新的3D模式设计得非常好的功能。 我们引入了一个基于卷积和递归神经网络（CNN和RNN）的组合的模型，用于学习特征和对RGB-D图像进行分类。 CNN层学习低层次的平移不变特征，然后给出多个固定的RNN，以组成更高阶的特征。 RNN可以组合进行卷积和合并，以进行有效的分层操作。 我们的主要结果是，即使是随机权重的RNN组成强大的功能。 我们的模型在一个标准的RGB-Dobject数据库上获得了最先进的性能，同时比类似的双层CNN架构更加准确和快速的训练和测试。

介绍：深度模态为复杂的一般物体检测问题[1]提供了有用的额外信息，因为深度信息是不变的着色或颜色变化，提供了几何形状，并允许从背景中分离。用RGB-Dimages识别的最新方法使用手工设计的特征，例如2维图像的SIFT [2]，3D点云的自旋图像[3]或特定的颜色，形状和几何特征[4,5]。我们引入了第一个卷积递归深度学习模型进行物体识别，可以从drawRGB-Dimages中得到。 比较前者3Dfeaturelearning方法[6,7]，我们的方法是快速的，不需要额外的输入通道，如表面法线，并获得最先进的成果检测家庭对象的任务。 图1概述了我们的方法。 培训和测试代码可在www.socher.org上获取。

I

我们的模型从原始RGB和深度图像开始，首先分别从中提取特征。每个模态首先被赋予一个单一的卷积神经网络层（CNN，[8]），它提供了有用的平移不变性的低层特征，如边，并允许对象的某些部分在一定程度上可变形。然后将汇集的滤波器响应给予递归神经网络（RNN，[9]），其可以学习组合特征和部分相互作用。 RNN通过具有相关权重和非线性的多层分层投影输入到较低维空间。

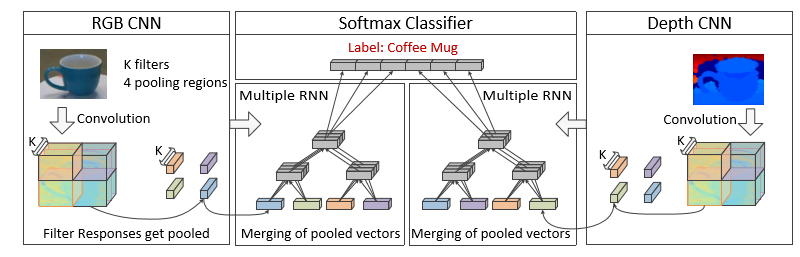


图1：我们的模型概述：单个CNN层从RGB和深度图像中提取低级特征。 这两个表示都被作为输入给一组具有随机权重的RNN。 每个RNN（每个模态大约100个）然后递归地将这些特征映射到较低维空间。 所有结果的相互关系形成了softmax分类器的最终特征向量。二进制，并（四）通过结构反向传播训练RNN [11,12]。 在本文中，我们通过在相同的输入上使用固定的树结构和多个RNN，并允许n元树，在这四个维度上扩展可能的基于RNN的体系结构的空间。 我们显示，由于CNN层，固定树结构不会损害性能，它可以使我们加快识别。 相似的工作[13,14]表明，NNN模型的性能可以提高许多特征。 将各种形式的分层组合RNN特性连接起来并赋予一个联合的softmax分类器。

我们首先简单地描述滤波器权重的无监督学习和它们的卷积以获得低级特征。接下来我们将详细介绍如何使用多个随机RNN来获取整个图像的高级特征。 然后，我们讨论相关的工作。 在我们的实验中，我们展示了不同模型的定量比较，分析了模型，并描述了Lai等人在RGB-D数据集上的结果。[2]。

卷积回归神经网络

2.1 UnsupervisedPre-trainingofCNNFilters

与在彩色图像的灰阶版本上训练的筛选器（右）相比，情况也是如此，尽管程度较低。

2.2 ASingleCNNLayer

CNN的主要思想是将输入图像上的滤波器进行卷积以提取特征。 我们的单层CNN与Jarrett等人提出的类似。 它由卷积，随后的确认和局部对比归一化（LCN）组成.LCN被计算神经科学所感知，用于对比特征图内的特征以及同一空间位置上的特征映射。我们用尺寸为dP的K个平方滤波器对每个尺寸（高度和宽度）的图像进行卷积，得到K个滤波器的响应，每个维度都是dI-dP + 1。然后，我们将它们与大小为d`的正方形区域和步幅 以获得具有等于r =（dI-d“）/ s + 1的宽度和高度的合并响应。因此，应用于一个图像的CNN层的输出X是K×r×r维3D矩阵。 我们将这个相同的程序分别应用于颜色和深度图像。

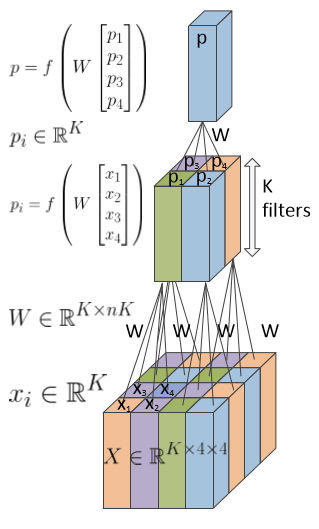
2.3 Fixed-TreeRecursiveNeuralNetworks

递归神经网络[19,9]的想法是通过在树结构中递归地应用相同的神经网络来学习分层特征表示。 在我们的例子中，树的叶节点是K维向量（CNN在所有K滤波器上重复的图像块的结果），并且有r2个。在我们以前的RNN工作中[9,10,20]，树结构取决于输入。虽然这允许更多的灵活性，但是我们发现对于与CNN层一起进行目标分类的任务来说，获得高性能是没有必要的。此外，对最优树的搜索大大减慢了方法，因为不容易并行搜索或利用大型矩阵产品的并行化。后者可以从GPU等新型多核硬件中获益匪浅。在这项工作中，我们关注固定树，我们可以设计平衡。以前的工作也只结合成对的向量。我们概括了我们的RNN体系结构，以允许每个层合并相邻向量的块而不是仅对部分。

我们从每个图像的3D矩阵X∈RK×r×r开始（列是K维的）。 我们定义了一些相邻的列向量，这些向量在不相关的向量p∈RK中。 在下面我们只使用方块。 块大小为K×b×b。 例如，如果我们合并b = 3的块中的矢量，则得到总尺寸为128×3×3的矢量列表（x1，...，x9）。 一般来说，我们在每个块中有许多矢量。

其中参数矩阵W∈RK×b2K，f是一个非线性如tanh。 我们省略了在下面的实验中不起作用的偏差项。 将等式（1）应用于X中具有相同权重W的所有向量块。通常，将会有（r / b）2许多父向量p，形成新的矩阵P1。 P1中的视图将再次被合并成块，就像矩阵X中的那些使用方程 1与相同的权重产生矩阵P2。 这个过程继续下去，直到只有一个无关的向量。 图3示出了具有4×4×4的容纳CNN输出的样本和具有4个孩子的块的RNN树结构。

以前的工作只使用一个RNN。 实际上，我们可以使用3D矩阵X作为多个RNN的输入。 每个RNN将输出一个K维向量。 在我们向前传播通过所有的RNN之后，我们把它们的输出连接到一个NK维向量，然后给予softmax分类器。 我们发现，即使是随机加权的RNNs也能产生高质量的特征向量，而不需要通过结构[11]来反推RNN的矩阵。 类似的结果已被发现在密切相关CNN的随机权重[16]。 在比较其他方法之前，我们简要回顾一下相关的工作。



2.4多个随机RNN

以前的工作只使用一个RNN。 实际上，我们可以使用3D矩阵X作为多个RNN的输入。 每个RNN将输出一个K维向量。 在我们向前传播通过所有的RNN之后，我们把它们的输出连接到一个NK维向量，然后给予softmax分类器。 我们发现，即使是随机加权的RNNs也能产生高质量的特征向量，而不需要通过结构[11]来反推RNN的矩阵。 类似的结果已被发现在密切相关CNN的随机权重[16]。 在比较其他方法之前，我们简要回顾一下相关的工作。

其他方法的比较在本节中，我们将我们的模型与文献中的相关模型进行比较。 表1列出了主要的准确性数字，并与公布的结果进行比较[2,5,6,7]。 Bo等人最近的工作。 文献[5]研究了各种特征的多个核描述符，包括三维形状，物体的物理尺寸，深度边缘，梯度，内核PCA，局部二进制模式等。 相比之下，我们所有的功能都是从原始颜色和深度图像以无监督的方式学习的。 Blum等人[6]