

CVDL第一次作业报告

李浩辰 1500012746

2018.4

1 简介

本次作业的目的是在给定的数据集上，对测试集的图片进行正确的场景分类。训练集包含80个类，数据集难度较大，因此最后采用了Resnet作为最终的神经网络架构[1]。Resnet是2016年所提出的深度卷积神经网络，其特点是梯度不易消失，因此可以有很深的层数，作者在不同的数据集上都取得了比较好的结果。本次采取了论文中所描述的Resnet-18结构。

除了Resnet之外，在寻找资料的过程中我还尝试了一种循环神经网络的结构[2]。可惜并未取得较好的结果，这部分会在下一节略作描述。

2 实现

2.1 预处理

输入的图像长宽比不同，大小也不同。所以将图像用双线性插值降采样到 224×224 大小的RGB三通道图像。

2.2 Resnet网络结构

论文中的标准Resnet-18网络，参数如下：

Table 1: Resnet网络结构			
conv1	conv2	conv3	conv4
conv $7 \times 7 \times 64$	res_block $3 \times 3 \times 64$	res_block $3 \times 3 \times 128$	res_block $3 \times 3 \times 256$
max-pool 3×3	res_block $3 \times 3 \times 64$	res_block $3 \times 3 \times 128$	res_block $3 \times 3 \times 256$
conv5	global_pool	fc.1	fc.2
res_block $3 \times 3 \times 512$	512	[512 \times 1000] with ReLU	[1000 \times 80]
res_block $3 \times 3 \times 512$			

每个conv层均包含一个conv，一个batch_norm和一个dropout。每个res_block包含两个conv层。当尺寸或通道数不对应时，会使用卷积层来做统一处理。

2.3 调参过程

起初模型不含有dropout，在经过了相似的训练之后最终收敛在了68%左右。之后希望通过添加dropout来减少模型的过拟合，综合考虑到Resnet的参数冗余较多，采用dropout可能是比

较好的方式。结果提高了约2%，但是训练时间也略微延长了一些，总的来说略有提升，但是并不明显，可能是需要表达能力更强的模型才能从根本解决问题。

2.4 尝试循环神经网络

首先将图像预处理成多个不同大小的图像。基本思路是对图像进行卷积，之后将卷积得到的特征与小尺寸的图像concat，之后对此继续做卷积，这里的卷积和之前共享参数，如此重复，最后将几次卷积得到的结果输入全连接层。

验证集的正确率收敛在了51%左右，参数共享的目的是期望能从大尺寸的图像和小尺寸的图像中发现共同的特征提取方式。从循环神经网络的性质来看也有从局部到整体的意味。但并未取得预期的效果。

3 测试

3.1 测试环境

Intel(R) Core(TM) i7-4720HQ 2.60GHz
Nvidia(R) GeForce GTX 970M
tensorflow-gpu 1.7

3.2 测试结果

Table 2: 验证集top3正确率和loss

Epoch	1	2	3	4	5
accuracy	0.420360	0.525586	0.585766	0.614955	0.631351
loss	3.0	2.6	2.4	2.25	2.18
Epoch	6	7	8	9	10
accuracy	0.659459	0.672613	0.683063	0.686486	0.696577
loss	2.07	2.02	1.97	1.95	1.90

10个epoch过后出现了过拟合的现象，因此停止了训练。最终的验证集正确率收敛在70%左右。

4 总结

Resnet在图像分类方面的效果十分显著，对比循环神经网络有十分显著的提高。Resnet对于Identity的使得类似Identity的映射变得更容易学习，多次使用batch_norm有利于梯度不消失，同时还可以部分解决过拟合的问题[1]。但是添加了dropout之后，网络过拟合依然比较严重，当训练集正确率达80%时，验证集的正确率向70%收敛，可能是网络的表达能力不够，考虑需要使用更深的网络才能达到更好的效果。

循环神经网络在此处的利用[2]可能不如它在音频等领域内表现优秀，图像可能并不存在某种意义上的先后关系，这与语音的时间还有文字的先后是很不一样的。

参考文献

- [1] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.
- [2] Pinheiro P H O, Collobert R. Recurrent Convolutional Neural Networks for Scene Labeling[C], International Conference on International Conference on Machine Learning. JMLR.org. 2014:I-82.