



EE369 课程大作业

M3DV项目——基于DenseNet的3D医学图像分类

陈昊 517021910970

2020年1月3日



上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

项目完成情况



- 最终Leaderboard上的分数： 70.997
- Leaderboard上显示的名字： 517021910970_陈昊
- 总共提交次数： 11
- 是否使用小号： 使用过小号增加提交次数验证模型
- 方法简述： 基于Keras/TensorFlow的
- 主要使用的代码框架： Keras/TensorFlow
- 模型大小（MB）： 1.1MB
- 亮点： 数据预处理、数据增广、集成学习
- 代码链接： <https://github.com/qqww1234/SJTU-M3DV>

问题描述



- 这是一个肺部结节的分类项目。本项目的目的是利用数据集训练并评估自己的模型，然后在测试数据集上验证自己的结果。
- 数据集是一系列的3D肺部结节的CT扫描。对于每个样本，包含：
- 一个大小为 $100 \times 100 \times 100$ ，已经经过预先的3D体素，其中心为肺部结节（病灶）
- 由医生宣布的结节分割mask
- 有某种方式得出的二分类结果，0/1
- 问题关键词：三维、二分类、过拟合

模型设计



- 使用DenseNet作为模型

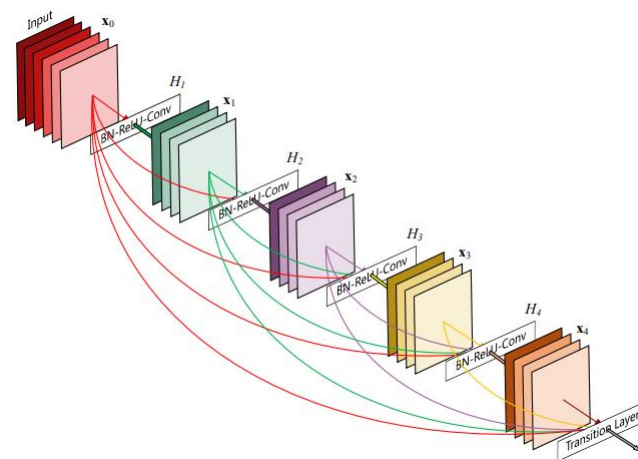
- 来源：参考资料https://github.com/titu1994/DenseNet/blob/master/densenet_fast.py

- 如右图是一个5层dense block的示意图，

每一层的输入都来自前面所有层的输出，
在减轻vanishing-gradient（**梯度消失**）方面有显著的效果。

本项目中采用depth=7，即nb_layer=1(这是由于扩大了batch_size而做出的让步)。

有关DenseNet的参考资料为论文Densely Connected Convolutional Networks
(<https://arxiv.org/pdf/1608.06993v3.pdf>)



性能分析



- 平均每层运算时间为7s, 在400代附近单个模型在测试集上的AUC可达到0.68左右,本地测试集准确率0.7+
- 采用双模型集成抑制过拟合, 最终配合一定的权重得到的最终模型在测试集上的AUC为0.70997
- 显卡使用: 1080 Ti
- 模型大小 (双模型) : 1.1MB

技巧设计



■ 1、数据预处理——有效结节位置的划定

由于原始数据集大小为 2×465 个 $100 \times 100 \times 100$ 的矩阵，但有效结节位置只占其中的一小部分，并且结节位置也已经由seg标定，因此首先根据结节位置提取 $32 \times 32 \times 32$ 的包含结节的信息，再按照0.8、0.2的权重分配voxel和seg，以此来进行数据的预处理，运算量减小的同时增大准确性。

■ 2、数据集的扩充及batch_size的选用

本项目用于训练的数据数量不多，使用小数据集训练卷积神经网络时最大的缺点就是很容易导致过拟合问题，因此采用data augmentation的方法，包括mixup和dropout等将数据集扩大了十倍。由于数据量增大，为了提升运算效率和减小震荡，经过反复实验选择batch_size=256的效果最优，可以在内存允许范围内最高效运算。因此也降低了模型中DenseBlock的层数。

技巧设计



▪ 3、earlystop

如何控制过拟合成为了项目中后期主要头疼的问题。往往Loss最低的模型过拟合非常严重，反而在本地数据集比较普通的中间批次结果更接近最优模型。因此通过设置earlystop可以便于寻找中间点，减轻过拟合带来的影响。甚至有一段时间过拟合问题无法解决时也采用过手动保存模型试验最佳保存点的方法。

▪ 4、集成学习抑制过拟合

将两个过拟合的模型按照一定的权重分配可以削弱两者的过拟合效应。利用这种方法可以增大单个模型的期望方差，相当于为两模型的参数均设置了惩罚因数，从而达到抑制过拟合的目的。在本项目中，采用的两个模型分别是做过数据增广（batch_size=256）和没做过数据增广（batch_size=16）的模型，二者误差相互独立。在本地数据集上二者均能达到0.7+的accuracy，但在验证集上最高只能做到0.68，将二者线性组合后AUC能够提升0.1-0.2，最终选择的较优权重为0.7与0.3。

技巧设计



- 被放弃使用的策略：

1、基于图像几何变换的data augmentation

开始进行数据增广时采用的是图像的几何变换，包括平移、翻折、旋转等，经过测试发现效果并不如意，分析可能原因是对于医学检测，病状的反转旋转等对于判定的结果可能有较大影响，而不是其他检测项目中的几何变换不改变性质，故最终放弃使用。

2、动态学习率

动态学习率是指在训练初期适当增大学习率，之后逐渐降低学习率的策略，本质上是提高效率以及逃离局部最优点的策略，但经过反复实验之后发现过小的学习率会加剧过拟合现象，因此放弃减小学习率。

讨论



- 是否可以使用更优的模型

在设想中简化模型也能很好地达到抗过拟合的效果，包括选择更优的优化器，也是基于这一点考虑选择了参数较少的DenseNet，但是缺乏对其他模型的了解导致可能没有选取到最优的方向。

- 是否有更好的测试集分类方法

本项目中采用的训练/测试集分类为HW2中使用过的train_test_split函数，对于其他的评估方法诸如交叉验证法等还没有去尝试，在数据集较小的情况下分类方式可能成为训练是否有效的重要因素。

问题



- 调参过程中并没有对损失函数进行调整，也没有尝试正则化等一些方法，整体上模型还是非常粗糙的，让我意识到自己离真正的专业还有不小的距离。
- 很多过程采用的是非常原始的方法，比如凭直觉调参、手动保存模型等，还需要对如何解决问题有更加系统的认知。

心得感悟



- 参加完这次的比赛让我受益匪浅。虽然过程比较痛苦（尤其是临近期末考试和大作业交织的那段时间），但是整个项目的过程中还是非常的专注和用心。对我个人而言，这次的学习过程是一次非常大的提升，也让我意识到自己还有很多不足。非常感谢老师、助教和同学们耐心的指导和帮助。

Thank You

