

Solution of the 'grt123' Team

May 5, 2017

1 引言

我们的模型是基于3D卷积神经网络的结节检测和癌症分类的统一框架。模型的输入是预处理过的3D肺部扫描，输出是可疑结节的边界框和患癌概率，概率用于计算损失值。该模型在可疑结节检测和全肺信息整合的过程中提供了可以解释的结果。我们相信对网络行为的进一步研究可能为了解癌症检测的病理机制提供有用的见解。解决方案可以分为三个模块，预处理，结核检测和癌症分类。在下面的内容中，我们将详细描述它们。

2 预处理

所有的原始数据首先被转换成Hounsfield units.

掩码计算

接下来我们计算肺部空间的掩码。对于每个切片(slice,序列中的一张二维图片)，使用高斯滤波器（size= 1像素）进行滤波，然后使用-600作为阈值进行二值化。所有小于30平方毫米的2D联通区域或偏心率大于0.99的部分均被移除。然后，我们找到所有二值3D矩阵中的所有3D连通区域，只保留那些不接触矩阵角的区域，其体积在0.68L到7.5L之间。

进一步分析。对于一个区域的每个切片，我们计算的面积（Area）和到切片中心的距离（MinDist）。如果一个面积大于6000平方毫米的区域的平均MinDist大于62毫米,则移除该区域。然后将其余的组件联合起来，形成基于图像强度的肺部区域掩码。

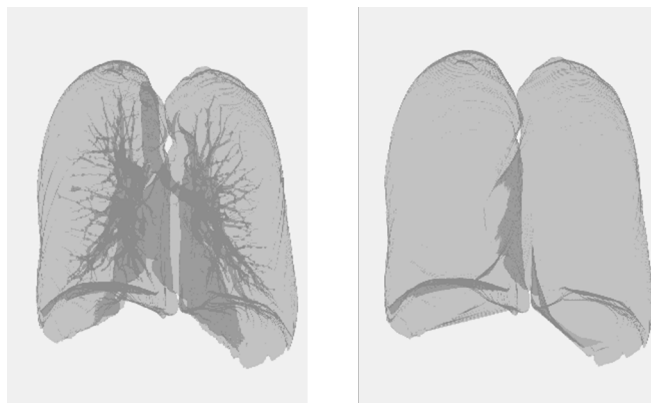


图1: 预处理中的肺部掩码计算. 左图是基于图像强度的原始掩码
右图是进行了封闭和扩充处理的,可用于对原始数据进行乘法运算的掩码

肺在某些情况下与外部空间联通,此时为保证预处理程序正常运行,需要将顶部的几张切片预先移除.

凸包& 扩张

对掩码进行腐蚀操作,直到最大的两个区域有相似的体积。将这两个区域扩大到原来的大小,并与原始掩码交,以此分别为两侧的肺生成掩码。对于一个肺部掩码的每个切片,如果增加的面积不超过50%,则2D二值图像被其凸包替换。掩码被进一步扩大10个体素来包括一些周围的空间。将两侧肺叶的掩码结合在一起就得到完整的掩码。

强度正则化

为了准备深度学习的数据,我们将Hounsfield unit中的数据转换为整型数据。首先用窗口 $[-1200,600]$ 对原始数据矩阵进行剪切,然后线性变换到 $[0,255]$ 。然后乘以上面得到的完整掩码,掩码外部用170填充。另外,对于最后一步中由膨胀产生的空间,所有大于210(通常是骨骼)的值被替换为170。填充值选择170是因为这是类似于组织的强度。

3 数据集

本文使用了两个肺部扫描数据集,LUNA16数据集(缩写为LUNA)和Kaggle stage 1 训练集(缩写为Kaggle1)。LUNA包括由放射科医师注释的结节标签,而Kaggle1仅有代表是否患癌的二值标签。

原始数据集有两个问题。首先，LUNA有一些非常小的注释结节，可能与癌症无关。其次，LUNA结节和Kaggle1结节之间存在一些显着的差异。Kaggle1存在两种特殊的结节，但在LUNA中罕见，包括大结节（大于60 mm）和主支气管上的结节。参见图2的例子。如果网络只在LUNA上进行训练，那么学习检测这些大的结节是很困难的。为了解决这两个问题，我们从LUNA的标注中移除小于6mm的结节，并手动标Kaggle1中的结节。

值得注意的是，作者对肺癌诊断没有专业知识，所以结节选择和手工注释可能会产生相当大的噪音。尽管如此，这些修改主要仅基于结节大小，并且在后面的实验中证明是有用的。更重要的是，下一阶段的模型（癌症分类）对错误检测是健壮的，从而减轻了对高度可靠标签的需求。

ea7373271a2441b5864df2053c0f5c3e

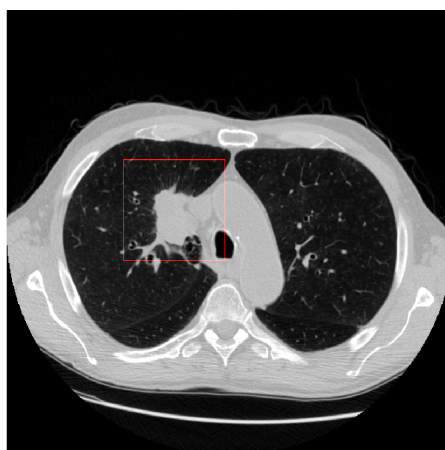


图2: 主支气管大结节的一个例子。

4 结节检测

我们设计了一个3D卷积神经网络来检测可疑的结节。它以一个3D版本的建议网络 (proposal net) 为基础, 针对具体任务进行了训练和网络结构的修改。与计算机视觉中涉及数十或数百个类别的通用对象检测不同, 在这个任务中只有两个类别, 所以预测建议直接作为检测结果, 而且没有采用在RCNN-type的检测器中使用的第二阶段分类器。

基于Patch输入的训练

对象检测模型通常采用基于图像的训练。在训练阶段，整个图像被用作网络的输入。但是，由于显存限制，这对于我们的3D ConvNet模型来说是不可行的。当肺部扫描的分辨率保持在一个合适的水平时，即使是单个样本，也会轻易耗尽主流GPU的最大显存。

为了克服这个问题，我们使用基于patch的输入。输入是从肺部扫描数据中裁剪出的一个 $128 \times 128 \times 128$ 的立方体。随机选择两种patch:首先，70%的输入被选择，以便它们至少包含一个结核目标。其次，30%的输入从肺部扫描中随机剪裁，可能不含结节,以确保覆盖足够的负样本。

如果一个patch超出了肺部扫描的范围，那么填充为170，与预处理相同。结节的目标不必位于patch的中心，但是距离边界要大于12个像素（除了一些太大的结节外）。

进行数据扩充以缓解过拟合问题。这些patch随机左右翻转，并以0.8-1.15之间的比例调整大小。还尝试了其他数据扩充方法，例如坐标轴交换和旋转。但它们没有显著的改进，因此最终没有采用。

主动样本平衡

尽管我们已经从LUNA中去除了非常小的结节，但是结节大小的分布仍然是高度不平衡的。小结节数量远大于大结节数量。如果使用均匀采样，网络将学习偏向于小结节。同时牺牲了大结节的准确性。这是我们不愿看到的现象,因为大结节通常比较小的结节更可能是癌症的指标。为了防止这个结果，我们增加了训练集中大结节的采样频率。特别的,大于30mm和40mm的结节的采样频率分别是其他结节的2倍和6倍。

难分样本挖掘(Hard negative mining)

[难分样本挖掘:当得到检测错误的patch时，会使用这个patch创建一个负样本，并把这个负样本添加到你的训练集中去。当你重新训练你的分类器后，分类器会表现的更好，并且会逐渐变得能分清原来难分的负样本.]

负样本比正样本要多得多，而且他们的分布非常不平衡。他们中的大多数很容易被网络的歧视，但其中也有一些负样本有带有和可疑结节相似的外观。作为目标检测的常用手段，通常采用难分样本挖掘来处理这个问题。我们在训练过程中使用一个简单的在线版本的难分样本挖掘。

第一步，我们使用网络处理patches，并获得输出"图"(我们原来把这个叫做输出向量/矩阵,因为到了输出这块,已经看出是什么图形了)，其中每个"像素"(其实就是维度)是分类置信度分数和回归项。第二步，随机选取N个负样本进入候选池。第三步，把这个池中的负样本根据分类信度得分降序排列，选前n个样本为难分样本。

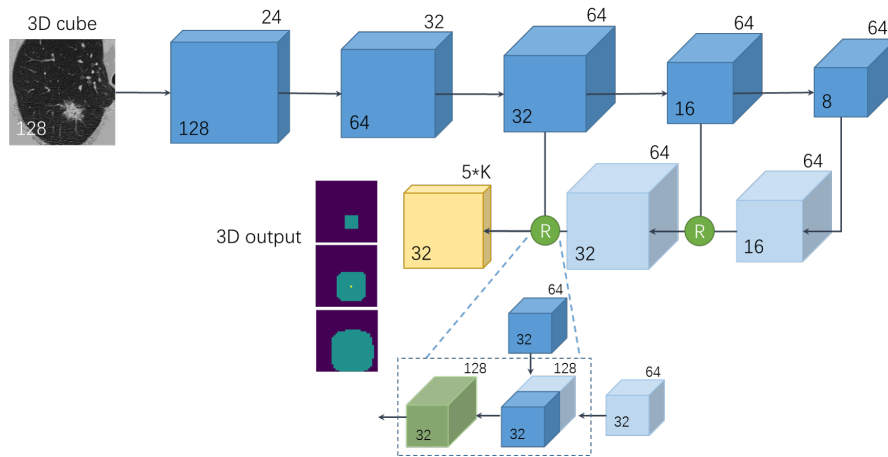


图3: 结节检测网络的结构

其他负面样本被丢弃，不参加损失计算。使用随机候选池可以减少负样本之间的相关性。通过调整候选池的大小和 n 的值，可以控制难分样本挖掘的强度。

网络结构

结节检测网络受U-Net的启发，其结构如图4所示。它由五组3d残差块,四个池化层,一组横向和反馈连接组成,最后的反馈层的特征用来计算输出,大小为 $32 \times 32 \times 32$.两个 $1 \times 1 \times 1$ 卷积用来获得分类分数和边界框回归项。

该网络有三个不同比例的锚点，分别对应10,30和60毫米。在最后的反馈层上用IoU确定类别, IoU大于0.5为正,小于0.02为负。其他像素在训练过程中被忽略。采用二元熵损失函数进行分类。回归项包含边界框回归标签 ($\Delta x/ x, \Delta y/ y, \Delta z/ z, \log (\Delta r)$)，并使用L1平滑损失。所以最终的标签图的形状是 $32 \times 32 \times 32 \times 5 \times 3$ 。

[IoU:简单来讲就是模型产生的目标窗口和原来标记窗口的交叠率。具体我们可以简单的理解为：即检测结果(DetectionResult)与 Ground Truth 的交集比上它们的并集]

测试阶段的完全卷积输出

网络训练后，整个肺部扫描作为输入，以获得所有可疑的结节检测。因为网络是完全卷积的，所以这是直截了当的。但是显存不足仍然是个问题,即使网络在测试阶段需要的内存少得多，仍然超过了我们GPU的最大显存。为了解决这个问题，我们将肺部扫描分成几个部分，分别进行处理，然后合并结果。请注意，我们保持这些分割大幅度重叠，以消除卷积计算过程中不需要的边界效应。

5 癌症分类

网络结构

训练结节检测网（简称N-Net）后，将其应用于每位患者的全部3D数据。根据他们在N-Net上的置信得分，挑选前五个预测结果。在训练阶段，预测结构采用softmax方式随机选出，而在测试阶段没有使用softmax。

我们提取了每个提议的N-Net的最后一个卷积层，它是128个特征的 $32 \times 32 \times 32$ 立方体，在图3中表示为绿色立方体。我们对每个预测的中心的 $2 \times 2 \times 2$ 体素进行最大池化，并将得到的128-D特征传递给两个连续的全连接层，预测该样本是癌症概率。我们假设最终的诊断是由这5个预测和一个假设的假结节（假如N-Net错过一些结节）独立完成的。预测损失定义为：

$$P = 1 - (1 - P_d) \prod_i (1 - P_i)$$

其中 P_d 是一个自由变量,代表从假想结节中获得的癌症可能性(图4).

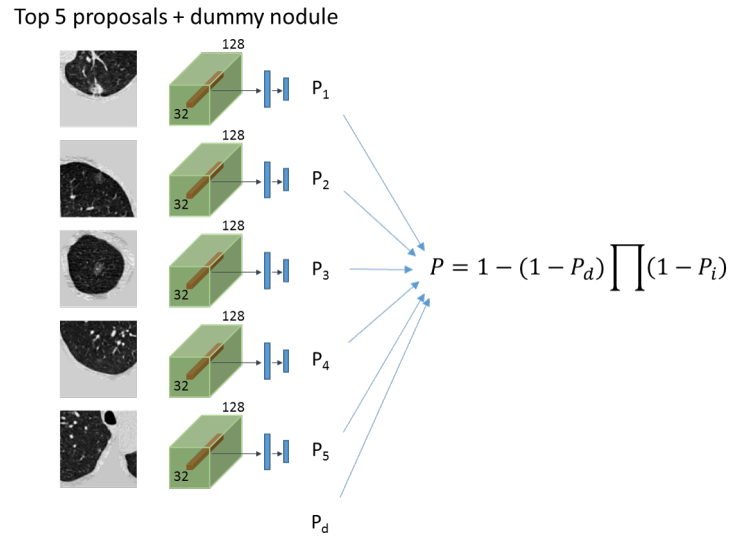


图4: 癌症分类网络的结构

增加额外的损失以促进训练. 如果 P_i 小于0.03并且在N-Net训练的时候被标记为结节 , 分类网络的损失值就增加 $-\ln(P_i)$.

Interleaving training

为了使N-Net具有可用性，我们在训练癌症分类网络（简称C-Net）时对其同时进行了训练。在每个时期，分别通过训练C-Net和N-Net来更新网络参数。卷积层在C-Net和N-Net之间共享。

第一阶段的分类任务只使用flip进行数据扩充扩充。在微调阶段使用了其他数据增强类型（flip, resize, swap, rotation）。

6 实现

我们采用随机梯度下降法（SGD）和常用的步进式学习率方案来训练网络。设置初始学习率，当损失停止减少时，学习率降低为原来的1/10。重复这个过程，直到网络收敛。batch大小根据模型规模和显存的大小来设置。

该模型是使用Pytorch实现的，这是一个非常灵活和高效的深度学习框架。使用了8个TITAN X GPU。