基于深层神经网络的脑肿瘤分割

**摘要**

本文提出了一种基于深度神经网络(DNNs)的全自动脑肿瘤分割方法。这种神经网络专门适用于磁共振图像中的恶性胶质瘤(包括低级别和高级别)。由于其本身性质，这些肿瘤可以出现在大脑的任何位置，具有各种各样的形状、大小和差异。这些理由激励我们寻找一种基于灵活、高容量DNN的高效机器学习方法。本文中，我们介绍了一些不同的模型选择，它们对于获得较高的性能来说是必要的。我们会在卷积神经网络(CNN)的基础上探索不同的结构，如图像数据适用的DNNs。

我们展示了一种新型的CNN架构，它有别于传统的计算机视觉。我们的CNN既利用了局部特征，同时也利用了全局上下文特征(global contextual features)。此外，与大多数CNN的传统方法不同，我们的网络最后一层使用一个能够40倍加速的全连接层的卷积实现。我们还设计了一个两阶段的训练程序，可以帮助我们解决肿瘤标签不平衡的问题。最后，我们探索一个级联结构，将一个基本CNN的输出作为后续CNN的一个附加信息源。根据2013年的BRATS测试数据集的结果显示，我们的结构比目前发布的最先进的技术提速了30倍以上。

1. **前言**

据估计，2015年在美国就有2.3万例新的脑癌病例被确诊。神经胶质瘤是最常见的脑肿瘤，有些神经胶质瘤(像低级别的)侵略性较低，患者预期寿命大约有好几年，有些更具有侵略性(即高级别的)，患者的预期寿命最多2年。

尽管手术是最常见的治疗脑瘤的方法，但是放疗和化疗只能减缓肿瘤的生长，而不能将肿瘤很好的移除。磁共振成像(MRI)提供了大脑图像的细节，是用于诊断脑瘤的最常见的尝试之一。更重要的是，磁共振图像中的脑肿瘤分割对肿瘤诊断、生长率预测和治疗规划都有很大的影响。

虽然分割像脑膜瘤这样的肿瘤很容易，但是分割神经胶质瘤和胶质母细胞瘤这样的肿瘤却很难。这些肿瘤(连同周围的水肿)经常扩散，对比度较差，并延伸出像触手一样的结构，使它们难以分割。大脑肿瘤难以分割的另一个原因是，它们可能出现在大脑的任何地方，各种各样的形状和大小。此外，与x射线计算机断层扫描(CT)扫描的图像不同，磁共振图像中的立体像素值的大小并不是标准化的。根据所使用的机器类型(1.5、3或7T)和采集协议(视野值、立体像素分辨率、梯度强度、b0值等)，在不同的医院里拍摄的同样的肿瘤细胞可能会有不同的灰度值。

健康的大脑通常由三种类型的组织构成：白质、灰质和脑脊液。脑肿瘤分割的目的是检测肿瘤区域的位置和延伸，即活跃的肿瘤组织(血管化或非血管化)、坏死组织和水肿(肿瘤附近肿胀)。这通过异常区域与正常组织相比较来识别。由于胶质母细胞瘤是侵入性肿瘤，它们的边界通常是模糊的，很难与健康的组织区分开来。常常使用如T1(自旋点阵弛豫)，T1-对比(T1C)，T2(自旋－自旋弛豫)，质子密度(PD)对比成像，扩散MRI(dMRI)，以及液体衰减翻转回复(FLAIR)脉冲序列等多种核磁共振成像。这些模式的对比给大脑组织不同的特征。

大多数自动脑肿瘤分割方法都使用了人为设计的特征(Farahani等人，2014；Menze等人，2014)这些方法实现了一个经典的机器学习流程，首先提取特征，然后把这些特征喂到一个不影响这些特征性质的分类器中去。另一种设计适应任务的特征的方法是直接从时域数据中学习复杂特征的层次结构。深度神经网络已经被证明擅长于学习这种特征的层次结构(Bengio等人，2013)。在这项研究中，我们运用这种方法来学习MRI模式中脑肿瘤分割的特征层次结构。

我们研究了几种CNNs的训练可能选择的方法，这些方法都是适用于图像数据的DNN。我们用完善的指标来说明他们的优点、缺点和性能。尽管CNNs首次出现是在20多年前 (LeCun等人，1998)，但由于在最近的图像网络的大规模视觉识别挑战(Krizhevsky等人，2012)中打破了记录，CNNs成为了计算机视觉社区的主流。虽然CNNs曾成功地应用于分割问题(Alvarez等人，2012年；Long等人，2015；Hariharan等人，2014；Ciresan等人，2012)，以前的大部分工作都集中在非医学的任务上，很多架构不太适用于医学图像或脑肿瘤分割。我们在BRATS 14研讨会上介绍了用卷积神经网络和其他两种方法进行脑肿瘤分割的初步研究。然而，这些结果是不完备的，需要更多的研究（详见第二节）。

在本文中，我们提出了一些特定的CNN架构来解决脑肿瘤的分割问题。我们的架构利用了CNN设计和训练方法的最新进展，比如max-out隐藏单元(Goodfellow等人，2013b)和Dropout正则化(Goodfellow等人，2014)。我们还调研了一些把肿瘤的局部形态和背景纳入考虑的架构。

许多机器学习方法有一个问题在于它们执行像素分类时不考虑标签的局部依赖性(即分割标签是条件独立于输入图像数据的)。考虑到这一点，可以使用结构化的输出方法，如条件随机字段(CRFs)，这种方法的计算成本较高。或者通过将最初的CNN的像素级别的概率估计作为附加的第二次DNN的额外输入，形成级联架构，可以对标签依赖性进行建模。由于卷积是一种高效的操作，所以这种方法比CRF实现要快得多。

我们将我们的实验分析聚焦在注释详尽的2013年MIC-CAI脑肿瘤分割(BRATS)挑战赛数据集(Farahani等人，2014年)上，该数据集使用良好定义的训练和测试，使我们可以直接定量地对各种方法进行比较。

我们的主要工作内容在以下四点：

1. 我们提出了一种完全自动化的方法，在2013年的排行榜上排名第二。
2. 我们的方法进行一次脑肿瘤的分割需要25秒到3分钟，这比大多数现有最先进的方法都要快一个数量级。
3. 我们的CNN采用了一种新型的双通道架构，它可以了解大脑的局部细节以及全局的上下文。我们还提出了一种两阶段训练步骤，这对于处理不平衡的标签分布非常关键。这些工作将在3.1.1和3.2中详细描述。
4. 我们采用了一种新型的级联架构，它高效且概念清晰，可以替代常见的结构化输出方法。有关这些模型的细节将在3.1.2节中给出。
5. **相关工作**

正如Menze等人(2014年)所指出的，在过去的几十年里，致力于自动脑肿瘤分割的出版物数量呈指数增长。这一发现不仅凸显了对自动脑肿瘤分割工具的需求，还表明该领域的研究仍在进行中。

脑肿瘤分割方法(尤其是用于核磁共振成像的)可以大致分为两类：基于生成模型和基于判别模型(Menze等人，2014；Bauer等人，2013；Angelini等人，2007)。

生成模型很大程度上依赖于特定领域的先验知识，即健康和肿瘤组织的外观。组织外观特征化有难度，而现有的生成模型通常将肿瘤作为一种与正常(或普通)大脑不同的形状或信号来辨识(Clark等人，1998)。通常情况下，先从一些健康大脑计算出来图谱或模板，这些方法依赖于3维磁共振图像经模板调整后的解剖模型 (Doyle等人，2013)。在Prastawa等人(2004)的研究中可以找到典型的大脑图像生成模型。考虑到ICBM的脑图谱，该方法将大脑与图谱相结合，并计算出健康组织(白质、灰质和脑脊液)的后验概率。肿瘤的区域是后概率低于一定阈值的位置。然后应用一个后处理步骤来确保良好的空间规则性。Prastawa等人(2003)也将大脑图像记录在图谱上，以获得异常情况的概率图，然后在图谱上初始化一个活动轮廓并迭代，直到后验概率的变化低于某个阈值。也有人提出其他一些活动轮廓的方法 (Khotanlou等人，2009；Cobzas等人，2007；Popuri等人，2012)，它们都依赖于左脑的对称特征或基于位置的特征。值得注意的是，由于在模板上校准肿瘤较大的大脑难度较大，一些方法同时进行配准和肿瘤分割(Kwon等人，2014；Parisot等人，2012)。

其他脑肿瘤分割方法采用判别模型。与生成模型不同的是，这些方法利用了很少的大脑解剖的先验知识，而主要依赖于（大量的）低层图像的特征提取，直接对这些特征与给定的立体像素的标签之间的关系进行建模。这些特征可能是原始输入像素值(Havaei等人，2014；Hamamci等人，2012)，邻域直方图(Kleesiek等人，2014；R.Meier等人，2014) ，像Gabor滤波器组这样的纹理特征(Subbanna等，2013；2014) 或者是基于图像的梯度、区域形状差异和对称分析(N.Tustison和Avants，2013)。经典的判别学习方法，如SVMs(Bauer等人，2011；Schmidt等人，2005；Lee等人，2005)和决策森林(Zikic等人，2012)。从2012、2013和2014年的MICCAI-BRATS挑战赛的结果来看，基于随机森林的方法是最准确的(Menze等人，2014；Gotz等人，2014；Kleesiek等人，2014)

判别模型的一个共同特点是实现了传统的依赖于人工设计特征的机器学习流程。在这些方法中，由于分类器与这些特征的性质是独立的，假定输入特征具有足够高的判别度，训练分类器才能将健康和非健康组织分开。基于人工设计特征的方法的一个难点是，为了在使用一些传统的机器学习方法时表现精确，它们常常需要计算大量的特征。这可能会使它们的计算速度慢，内存成本高。更高效的技术可以使用降维或特征选择的方法来使用更少的特征，但是特征数量的减少通常是以降低精度的成本为代价的。

从本质上讲，许多人工设计特征利用了泛型边缘相关的信息，而没有对脑肿瘤方面进行特异性适应。理想情况下，我们希望能够将一些特征组合到更高级的、适应任务的表现形式中。最近，初步调查显示，使用CNNs进行脑肿瘤分割是一种非常有前景的方法(参见BRATS 2014年挑战研讨会文献Davy等人(2014)；Zikic等人(2014)；Urban等人(2014))。这三种方法都将3维图像划分为2维(Davy等人，2014；Zikic等人，2014)或3维补丁(Urban等人，2014)，并且训练一个CNN来预测它的中心像素类。Urban等人(2014)和Zikic等人(2014)实现了一个很常见的的CNN，由一系列卷积层组成，每层之间是非线性激活函数，采用soft-max作为输出层。我们本文中工作扩展了我们在Davy等人(2014年)中提出的初步结果，使用了一种双通道架构作为本文的一个组成部分。

在计算机视觉中，基于CNN分割的模型通常被应用于自然场景标记。在这些任务中，模型的输入是颜色图像的一个RGB通道补丁。Pinheiro和Collobert (2014)的工作使用了一个基础的CNN来预测每个像素，并将预测结果作为第二个CNN模型的输入来进一步提高预测。其他的工作(Farabet等人，2013)涉及到用不同的CNNs处理不同分辨率的图像。最终的每个像素类预测是通过从所有CNNs中学习的信息集成来实现的。为了实现平滑的分割，这些预测使用了一个更全局的超像素图像分割。和我们的工作相似，其他最近的工作也利用了网络的最后一层卷积运算来扩展传统的CNN架构来实现语义场景的分割(Long等人，2015)。一般来说，在医学成像领域使用CNNs进行分割的工作相对较少。然而，最近Huang和Jain (2013)的一些著名的研究已经用CNNs来预测电子显微镜图像中神经组织的边界。这里我们探讨的方法与前面讨论的各种方法有相似之处，但是是在脑肿瘤分割的背景下。

1. **我们的卷积神经网络方法**

由于BRATS数据集的大脑在第三维中缺乏分辨率，因此我们考虑从轴向视图中对切片进行分割。因此，我们的模型依次处理每一个二维轴向图像(切片)，其中每个像素都与不同的图像模式T1，T2，T1C和FLAIR相关联。和大多数基于CNN分割的模型一样(Pinheiro和Pinheiro，2014；Farabet等人，2013)，我们的方法通过处理以像素为中心的M×M这一小部分来预测一个像素的类。CNN模型的输入X是一个有多种形式的2维M×M区域。

用于构建CNN架构的主要组成部分是卷积层。可以将多个层堆叠在一起，形成一个层次结构。每个层都可以被理解为将其前一层的特性提取到它所连接的层次结构中。一个卷积层可以作为输入层的输入，并输出一些输出平面或特征图。每一个特征图都可以被看作是一个拓扑排列的地图，它是一个特定的空间局部非线性特征提取器(学习的参数)，在一个滑动窗口中对输入平面的每个空间区域都是相同的。在第一个卷积层的情况下，单个的输入平面对应于不同的MRI模式(在典型的计算机视觉应用中，单个的输入平面对应于红、绿、蓝的颜色通道)。在随后的层中，输入层通常由前一层的特征映射组成。

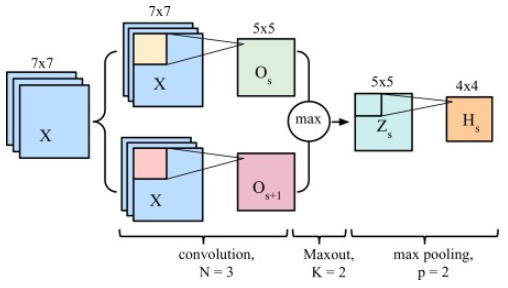


图1 一个单独的卷积层块，显示单一特征映射的计算。

输入部分(这里是7×7)与一系列的内核(这里是3×3)连接，然后是max-out和max-池化。

计算卷积层的一个特征图(见图1)由以下三个步骤组成：

1. 内核卷积(过滤器)：每个特征图都与一个内核(或多个，在Max-out的情况下)相关联。特性图计算如下：

其中是第个输入通道，是那个信道的子核，是卷积运算，是一个偏置项。换句话说，为每个特征图执行的仿射操作是个不同的二维卷积滤波器的应用的总和(一个输入通道/通道)，加上一个偏置项，每一个产生的空间位置都增加了像素。尽管这个操作的输入是一个3维M×M×R的张量，但是空间拓扑在原始脑容量的x-y轴平面上是二维的。

而传统的图像特征提取方法依赖于一个固定的方法(有时像线性卷积的形式如Gabor滤波器组), 卷积神经网络成功的关键在于他们能够学习单个特征图的权重和偏差，从而产生数据驱动的、自定义的、特定于任务的密集特征提取器。这些参数是通过随机梯度下降法对与误分率相关的代理损失函数进行调整的，通过反向传播算法(Rumelhart等人，1988)有效地计算了梯度。

通过卷积运算，必须对边界像素的处理给予特别的关注。在整个体系结构中，我们采用了所谓的“有效模式卷积”，这意味着小于的像素位置的滤波器响应不会被计算到。一个N×N滤波器与一个M×M输入区域进行卷积，将会产生一个Q×Q的输出，即Q=M－N+1。在图1中，M=7，N=3，因此Q=5。注意，内核的大小(空间宽度和高度)必须是由用户指定的超参数。

1. 非线性激活函数：为了获得输入的非线性变换的特征，应用于核函数卷积的结果。这种非线性有多种选择，比如s型、双曲正切和矫正线性函数(Jarrett等人，2009)，(Glorot等人，2011)。

最近，Goodfellow等人(2013b)提出了一种Max-out非线性特征，它被证明在建模方面特别有效。Max-out特征与多个内核相关。这意味着每个映射到的映射都与K个特征图相关联：。注意图1中，Max-out图与K=2的特征图相关联。Max-out特征相当于将最大的特征映射到特征图上，分别对应于每个空间位置：

其中是空间位置。Max-out特征相当于使用一个凸激活函数，但其形状是自适应的，取决于内核所取的值。

1. Max池化：这个操作包括在每个特征图的子窗中使用最大特征(神经元)值。这可以被形式化为：

其中决定最大的池化窗口大小。子窗口可以重叠或不重叠(图1显示了重叠的结构)。使用max池化操作缩小了特征映射的大小。这是由池的大小和跨步超参数控制的，对应于池的子窗口的水平和垂直增量。让S成为跨步值，Q×Q是在合并之前的特性图的形状。这个集合的输出的输出大小为D×D，D=(Q-p)/S+1。在图1中，由于Q=5、p=2、S=1，max池化操作结果为D=4输出特征图。这种操作的动机是为局部翻译引入不变性。这个子抽样过程在其他应用程序中被发现是有效的(Krizhevsky等人，2012)。

卷积网络有能力提取出越来越复杂的特征，这使得它们非常有吸引力。通过将卷积层的输出特性映射作为后续卷积层的输入通道来完成。

从神经网络的角度来看，特征图对应的是一层隐藏的单元或神经元。特别地，一个特征图中的每个坐标对应于一个单独的神经元，因为它的接收域的大小对应于内核的大小。内核的值也代表了层的神经元和前一层神经元之间的连接的权重。在实践中经常发现，学习的内核类似于边缘探测器，每个内核都被调到不同的空间频率、规模和方向，这与训练数据的统计数据是相对应的。

最后，为了对分割标签进行预测，我们将最后一个卷积隐藏层连接到一个卷积输出层，然后是一个非线性(即不执行共享)。需要注意的是，由于目的是划分，传统的CNN不会产生一个有效的测试时间，因为输出层通常是完全连接的。我们通过在末端使用卷积进行有效的实现，在测试时，整个大脑的预测将会快45倍。卷积使用的内核和不同的细分标签一样(在我们的例子中是5)。因此，每一个内核都充当了一个分割标签的最终检测器。我们使用软的非线性，将核卷积的结果规范化成一个多的分布在标签上。

3.1 体系结构

到目前为止，我们对CNNs的描述提出了一种简单的架构，对应于单个的卷积层。这种配置是计算机视觉文献中最常用的架构。但是，其他的体系结构可能更适合于即将到来的任务。

在这个工作中，我们通过使用来自不同层的特征映射的连接来探索各种体系结构，这是构成CNNs的另一个操作。该操作允许我们构建具有多个计算路径的体系结构，这些路径可以为不同的目的提供服务。现在，我们将描述这两种类型的体系结构。

3.1.1 双通道架构

这个架构由两个流组成:一个较小的7×7个接收域的通道，另一个有更大的13×13个接收域。我们将这些流分别称为局部路径和全局路径。我们选择这种架构的动机是希望像素的标签的预测受到两个方面的影响：一个像素周围区域的视觉细节和它的更大的背景，也就是大致的区域。

图2展示了完整的体系结构及其细节。我们把这个架构称为TwoPathCNN。为了让两个通道的顶部隐藏层连接起来，我们使用了两个层作为本地路径，在第二个层上有3个3个内核。虽然这意味着在每个路径的顶层的特征的有效接受域是相同的，但全局路径的参数化更直接和灵活地在同一区域内建模。然后将这两个路径的特征映射连接到输出层。