强调

* •

我们提出了一个快速的MRI海马分割软件，已经在大型队列中验证。

* •

该方法使用深度学习（3D ConvNets）并从其他现有（已实施）知识中学习。

* •

对于培训，我们使用了强大的扩增方案，将真实数据和合成数据混合使用。

* •

该方法是可推广的，并且可以应用于从标签充满或近似的许多医学数据问题中学习特征。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

抽象

由于其在人类大脑中的重要作用，海马是神经科学研究特别有趣的目标。在大型人类队列研究中，双侧海马结构经常被识别和测量，以深入了解人类行为或感兴趣的神经精神障碍的基因组变异性。自动分割使用各种算法执行，FreeSurfer是一种流行的选项。在这篇手稿中，我们提出了一种使用深度学习外观模型来分割双侧海马的方法。深卷积神经网络（ConvNets）近年来取得了巨大成功，因为它们能够从大量训练数据中学习有意义的特征。我们的方法依赖于以下关键的新颖事物：（i）我们使用来自多个队列的广泛和可变的训练集（ii）我们的训练标签部分来自FreeSurfer算法的输出，并且（iii）我们包含合成数据并使用强大的数据增强方案。我们的方法被证明是稳健的，并且具有快速推断（每个学科总共<30s），在线提供训练有素的模型（<https://github.com/bthyreau/hippodeep>）。我们描述了说明性的结果，并展示了与FreeSurfer进行广泛的定性和定量队列范围比较。我们的工作表明，深层神经网络方法可以很容易地编码，甚至改进现有的解剖知识，即使这些知识以算法形式存在。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

1.介绍

海马在人脑中发挥的重要作用使其成为神经影像学研究的一个特别重要的目标。海马因其功能参与记忆机制而众所周知，并且还通过其结构直接与记忆相关的疾病如阿尔茨海默氏病相关联（[Bobinski等，2000](javascript:void(0);)，[Jack等1999年](javascript:void(0);)，[Laakso等人，1996年](javascript:void(0);)，[Reiman等人，1998年](javascript:void(0);)，[Schuff等人，2008年](javascript:void(0);)，[西米奇等人，1997年](javascript:void(0);)）和PTSD（[Bonne等人，2001](javascript:void(0);)，[Gilbertson等人，2002年](javascript:void(0);)，[Gurvits等人，1996](javascript:void(0);)，[Pederson等人，2004](javascript:void(0);)，[Shin等人，2004](javascript:void(0);) ; 但见[Landré等人，2010年](javascript:void(0);)为阴性结果）或间接在其他疾病如精神分裂症（[Koolschijn等，2010](javascript:void(0);)），强迫症（[Boedhoe等，2016](javascript:void(0);)），抑郁症（[Bremner et al。，2000](javascript:void(0);)，[Janssen等，2007](javascript:void(0);)，[Vythilingam等，2002](javascript:void(0);)），痴呆（[Wachinger等，2016](javascript:void(0);)），甚至自闭症（[Aylward等，1999](javascript:void(0);)，[Sparks等，2002](javascript:void(0);) ;但参见[Piven等， ，1998年](javascript:void(0);)）。因此，通过提供关键信息，海马结构的定量分析是大规模基于MRI的精神病人群研究的重要方面。作为一个单一的，但优秀的实施例中，ENIGMA财团-第一非常大的[ *Ñ*  > 30000]在大脑基因变异的大型分析-都把海马体积作为选择的内表来执行基因组变异性（共享研究[Stein等人。，2012](javascript:void(0);)）。

在大群体的情况下，必须从自动MR图像分析算法获得海马体积。实践中最流行的工具之一是FreeSurfer（[Fischl，2012](javascript:void(0);)），这是一个丰富的脑部MRI分析平台，可输出对皮质表面研究特别有用的全套脑指标。FreeSurfer海马体积结果已被证明是相当稳健的（[Morey等，2010](javascript:void(0);)，[Sánchez-Benavides等，2010](javascript:void(0);)），但一个缺点是运行该程序非常耗时且耗费资源，这使得它很昂贵即使在今天，也可以在大型队列中运行。

尽管它们在大型人群中应用较少，但其他许多方法（[Carmichael等，2005](javascript:void(0);)，[Collins和Pruessner，2010](javascript:void(0);)，[Coupé等，2011](javascript:void(0);)，[Tangaro等，2014](javascript:void(0);)））或工具，如FSL First （[Patenaude et al。，2011](javascript:void(0);)）也已经开发出来。大多数工具依赖于脑图谱的初始配准，然后是几何海马模型拟合，在空间匹配算法和后处理方面具有不同程度的复杂性。被称为多重图谱分割的进一步改进（[Klein等人，2005](javascript:void(0);)，[Rohlfing等人，2003](javascript:void(0);)）依赖多个图谱集的存在来改善通过标记传播的分割（[Khan等，2011](javascript:void(0);)）。请参阅[Iglesias和Sabuncu（2015](javascript:void(0);)）的评论。这已被应用于7特斯拉的海马分割（[Kim et al。，2013](javascript:void(0);)），其中易于使用的实现可从[Pipitone等人获得。（2014年](javascript:void(0);)）。一个限制是它需要用户提供足够的地图集。至关重要的是，所有这些方法在很大程度上取决于具有良好的注册过程（尽管有一些缓解这个问题的建议（[Jia et al。，2012](javascript:void(0);)））以匹配图像和地图集。

在这里我们提出一种海马分割方法，它不会尝试翘曲图集。相反，它依赖于从现有数据中学习到的“深”海马外观模型。该方法应用卷积神经网络（ConvNet）从固定大小的输入区域框识别和分割海马。ConvNets在机器视觉领域取得了巨大成功，并被认为是目标识别的最新技术。这是由于他们能够通过从大型训练数据集中学习一组卷积滤波器来处理图像并提取越来越抽象的特征。简而言之，输入图像通过一系列图层，其中在每个图层中，多个卷积核被应用于输入要素以产生下一图层的特征。[Bottou，2010](javascript:void(0);)）。ConvNets提供了其他优点，如相对较少的参数数量，使得它们更容易训练，过度配合较少，并且具有快速推理。

尽管在[Albarqouni等人的研究中](javascript:void(0);)成功地使用了医学影像分析的深度学习方法[。（2016年](javascript:void(0);)），[Ciompi等人（2015](javascript:void(0);)），[Ciresan等人，2012](javascript:void(0);)，[Ciresan等人，2013](javascript:void(0);)），[Kallenberg等人 （2016](javascript:void(0);)），[Liskowski和Krawiec（2016](javascript:void(0);)），[Liu和Shen（2014](javascript:void(0);)），[Pereira等人 （2016a](javascript:void(0);)），[Prasoon等人 （2013](javascript:void(0);)），[Wang等人 （2014年](javascript:void(0);)）和[谢等人。（2016](javascript:void(0);)），其对于由3D体积构成的数据的具体使用仅在最近才变得流行，这可能是由于计算困难（3D卷积的高效，并行计算），需要大量内存的图像大小以及训练数据可用性等原因。ConvNets已成功应用于心脏成像（[Margeta等，2015](javascript:void(0);)，[Poudel等，2016](javascript:void(0);)），肾（[Thong等，2016](javascript:void(0);)），或脑（[Bao and Chung，2016](javascript:void(0);)，[Chen et al。，2016b](javascript:void(0);)，[Kleesiek等人，2016](javascript:void(0);)，[Moeskops等人，2016](javascript:void(0);)，[Zhang等人，2015](javascript:void(0);)）⁠自动识别结构。此外，ConvNets还帮助定位脑微出血（[Dou等，2016](javascript:void(0);)），脑节或病理性脑损伤（[Brosch等，2016](javascript:void(0);)，[Ghafoorian等，2016](javascript:void(0);)，[Kamnitsas等，2017](javascript:void(0);)，[Valverde等，2017](javascript:void(0);)），或肿瘤（[Davy等。，2014](javascript:void(0);)，[Pereira等，2016b](javascript:void(0);)）⁠。

在这些学习方法中，为数据的一个子集提供了一个感兴趣的目标，并且ConvNet必须概括一个测试集。但是，在实践中，经常出现两个问题。首先，培训材料通常依赖于手工标记的基本事实，这限制了总数据集的大小。其次，训练和测试集合通常具有高度相似的MR对比度，并且建议在每个新数据集上重新训练ConvNet。虽然这可以对方法的行为进行良好的评估，但它将其应用严格限制于现实世界的设置，其中一种方法继续适用于新的MR序列参数是优选的。在这个手稿中，生成一个可以在真实世界中直接使用的模型是一个主要目标。

我们提出了一个神经网络模型，能够在标准硬件上在不到30秒内（主要是由于初始化和近似预注册）分割单个T1脑图像的双侧海马。它的核心是一个深度3D卷积网络，训练来自大量真实和合成示例。我们不是从头开始，而是依靠FreeSurfer和现有FreeSurfer标记的在线数据集的存在，将其作为一个巨大的知识来源。这使我们能够快速大幅度增加培训样本的数量。但是，由于这会限制FreeSurfer掩模的精度，训练集还包含大量高精度，高SNR（信噪比）合成样本。训练实例的混合，结合非线性几何数据增强过程，倾向于使网络正常化，从而使其在解剖变异性和MRI序列变化性方面稳健。少量的参数迫使网络学习有意义的特征，这些调和可能相互矛盾（由于不精确的标签）训练样本。

在这篇论文中，转移学习是指从复杂算法的输出中对模型进行训练。在深度学习社区中，转移学习也是在较小数据集的更具体的领域上对预先训练过的大型和更通用数据集的模型进行微调的同义词。作为一个突出的例子，[Esteva et al。（2017年](javascript:void(0);)）最近在黑色素瘤图像的大型数据集上对ImageNet预训练网络进行了微调。他们的模型通过单一图像在皮肤癌预测中达到了皮肤科医师级别的准确性。另一种更简单的方法是使用大型预先训练的模型作为通用特征提取器并直接对其进行分类。遵循这种方法[Shie et al。（2015年](javascript:void(0);)）以90％的准确度从相应的图片中诊断出中耳炎。第三种方法将来自大型数据集的无监督训练和更小的标记集合上的有监督训练相结合，这是[Van Opbroek等人](javascript:void(0);)使用的一种设置[。（2015](javascript:void(0);)）⁠训练加权SVM分类器对多个MRI图像对比度进行推广，超出标注数据中的一个。[Shin等人对](javascript:void(0);)这些不同的ConvNet培训方案进行了比较[。（2016](javascript:void(0);)）关于胸部二维图像和肺间质图像异常检测的两项任务。他们比较了完全监督训练，无监督训练和监督微调以及使用多种流行体系结构（AlexNet，VGG，GoogLeNet等）对自然图像进行预训练的模型进行了微调，他们报告了从ImageNet到计算机辅助诊断始终有益。

这些模型正在跨图像域或整个对比度传递知识。但是，我们在这里跨软件传输知识。到目前为止，这还不是一种普遍的技术，可能是因为它仅在原始算法存在限制时才有用，例如速度，成本或针对特定用途的适用性。事实上，我们必须使用增强和模拟来补充输出。过去，合成数据一直是评估甚至训练机器学习模型的便捷方式。虚拟环境（例如，[You et al。（2017](javascript:void(0);)））一直是计算机视觉模型的来源。此外，ConvNet和真实数据现在正用于改善模拟样本的真实性（[Shrivastava等，2016](javascript:void(0);)，[You等，2017](javascript:void(0);)）。在医疗领域，[Itu等人 （2016年](javascript:void(0);)）使用机器学习系统对完全合成数据进行训练，根据基于物理学模型计算的目标，从冠状动脉CT扫描中估计分数流量储备。[施奈德等人。（2015年](javascript:void(0);)）使用转向过滤器和随机森林以高分辨率体积分割血管。为了达到较高的准确性和鲁棒性，他们依赖生理合理模型（[Schneider et al。，2012](javascript:void(0);)）生成的动脉树模拟图像作为分类器训练和测试输入的来源。合成目标也可以与真实数据混合使用。为了追踪X射线图像上的超声换能器，[Heimann等人 （2014](javascript:void(0);)）通过使用在合成图像上训练的鉴别分类器，在将换能器外观混合在真实的空X射线图像上的情况下实现了故障率的显着降低。在本文中，我们的合成数据是通过在几何和强度方面改变高质量模板而创建的。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.材料和方法

我们训练了海马外观的卷积神经网络模型。首先，我们描述模型架构及其训练过程。然后我们详细介绍我们的训练集和我们的数据增强方案。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.1。分割模型

我们的方法首先使输入近似注册到一个固定大小的训练感兴趣区域，然后将海马区域馈送到我们的ConvNet（在右海马的情况下左右翻转），其输出概率分割图。该地图投影回原生空间，并可选择设置阈值。海马的体积可以在原生空间或模板空间中计算，并相应地缩放。[图1](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952865/gr1.jpg)描绘了整个过程。

[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952865/gr1.jpg)

图。1

我们模型的一般架构。预处理和后处理调用外部工具。ConvNet在中间部分使用标记为：特征图数量（高度），数据大小（宽度），图层类型的标准框描述进行描述。灰色框包含无法训练的参数。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952865/gr1.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952874/gr1_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr1.jpg)

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.1.1。预处理

对于初次注册，我们重复使用标准的神经影像工具。具体来说，我们有脚本可以调用FSL FLIRT（[Jenkinson et al。，2002](javascript:void(0);)，[Jenkinson and Smith，2001](javascript:void(0);)）或ANT（[Avants et al。，2011](javascript:void(0);)，[Avants et al。，2009](javascript:void(0);)）。一个优点是，如果一个算法的注册失败（很容易从启发式中检测到，例如异常注册雅可比矩阵），则可以使用其他算法而不会产生副作用。它们的运行时性能大致相同。为了获得最大的速度和精度，我们使用两步注册过程：下采样输入首先严格注册到一个低分辨率（4毫米），全头模板。然后使用仿射变换将目标区域（1毫米）进一步注册到目标边界框。盒子包含双侧海马加上1毫米各向同性分辨率的附加背景。该盒子被进一步拆分，其右侧翻转，这简化了网络，只学习左侧。

在进入卷积网络之前，输入图像被强度归一化以补偿T1对比MR信号的大范围。由于空间覆盖率较小，因此无需运行MRI强度偏差校正算法。

卷积网络拓扑[图1](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952865/gr1.jpg)的主要部分描述了我们的分割深度卷积神经网络的体系结构。它由一个编码和定位器网络组成，连接到一个掩模生成网络，然后是一个细化网络。它的设计非常简单，在其他医学环境中使用拓扑接近U-Net来分割切片（[Çiçeket al。，2016](javascript:void(0);)，[Ronneberger et al。，2015](javascript:void(0);)）。我们使用体积大的内核，这些内核在脊柱分割任务中表现出比基于切片的内核更好的工作（[Chen et al。，2016a](javascript:void(0);)）。

首先，通过沿着每个维度的三个连续的一维卷积滤波器来对输入进行滤波。这往往会突出显示包含感兴趣特征的边缘，并且由于已经表明，1D滤波器的层数可以以较低的存储成本近似于二维等值（[Szegedy et al。，2015](javascript:void(0);)），因此将该想法应用于三个维度。一个完整的3D卷积层随后使用48个特征图，这被选为代表性容量和内存消耗之间可接受的折衷。下一部分首先使用减少空间大小的maxpool层（从而对更多空间不变特征敏感），然后是两层3D卷积滤波器。我们使用剩余连接（[He et al。，2016）](javascript:void(0);)），已经证明它具有适当的拓扑结构（[Zagoruyko和Komodakis，2016](javascript:void(0);)）[，](javascript:void(0);)可以提高训练收敛速度（[Szegedy et al。，2016](javascript:void(0);)）。我们使用的所有卷积核的大小为3.再次重复maxpool-ResNet模块，这将生成48个特征映射，其空间维度比原始数据小四倍。

在这一点上，相关信息被编码，使得具有3D卷积的两个放大层可以学习生成海马概率图。由于分割是通过升级来自低空间分辨率的数据创建的，因此我们添加了一个细化网络，该网络经过训练可重用初始图层中的特征以改进描绘。卷积滤波器在空间上不变，但我们希望只关注海马周围的区域; 因此，所需特征的初始映射在直接海马区域之外被设置为零（其通过平滑然后阈值化当前概率掩模来定义，概念上相当于执行分割掩模的体素扩张）。最终的结果是一个概率分割，

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.1.2。损失函数和优化参数

该网络的参数通过监督训练进行优化。尽管我们的目标原则上是二元的（无论是否海马组织），但是我们在损失函数中使用连续的结果，这允许标签的不确定性。这种概率分割也可以解释为每个体素内的部分体积。我们使用了损失：

学公式

其中*t*是目标标签，*y 1*和*y 2*是两个网络输出（在[图1中](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952865/gr1.jpg)表示为output\_1和output\_2 ），MSE是体素方均方算子，*α*（α）是一个常数标量。也就是说，损失函数使用细化和非细化分割输出错误的加权和。阿尔法的价值在整个训练过程中演变; 在预热阶段，α被设置为0，然后在大多数训练中α被设置为0.75，然后在最后阶段α被设置为1。

这些训练使用随着Nesterov动量的随机梯度下降进行，15个小批量的小批量，这是我们可以在我们的GPU内存中最大的。这个看似小批量的产品并不妨碍融合。事实上，已经表明，小批量规模倾向于避免在当地最低限度上解决（[Keskar等，2016](javascript:void(0);)）。然而，在训练一开始就有一套当地最低标准。为了避免这个初始障碍，网络接受了单个时代的培训，以产生平滑的海马图像，而不是每个样本的真实面具。预热时的学习率为0.001，然后设定为0.01。时代的总数是500。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.2。培训数据

作为一个监督学习系统，训练数据由一对图像（一个海马区域图像，一个相应的蒙版图像）组成，并且该模型通过分析数千个提供为训练的这样的对来学习。我们的训练集包含两种类型的输入：首先，使用FreeSurfer分割的四个队列中的2500个对象数据（×2边）作为学习目标。另外，数据集大小通过下面描述的增强过程增加三倍。增强是指对真实数据（及其目标）应用特定变换以人为创建更多的训练数据，这有助于学习。其次，大量的合成训练样本（10,000）是使用相同的增强方法从单一来源的变体中产生的，如下所述。整套训练集有25,000个数据点。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.2.1。队列图像

包含在我们训练集中的真实参与者的2500 T1 MR图像（3D数据）来自多个MRI队列的子集。由于我们的尺寸不受任何昂贵的手动标签处理的限制，训练数据集可能是无限的。因此，我们从以下四个队列中选择了任意和不等数量的受试者。尽管如此，我们仍然保留大量数据，以便稍后进行详细验证。在数量上，我们的训练集包括：

**HCP**。来自Human Connectome项目的350个随机对象（[Glasser等，2016](javascript:void(0);)），“900名受试者无限制”的数据发布。这些数据包含在3T扫描仪上以0.7 mm的原始各向同性分辨率采集的高质量，高对比度图像（[Van Essen et al。，2012](javascript:void(0);)）。

**AOBA**。来自我们自己的AOBA数据库的900个随机科目。该数据集（[Sato et al。，2003](javascript:void(0);)）是最初的1547名日本16〜79岁正常人的终身队列，在GE-Yokogawa 0.5 Tesla MR扫描仪上以1×1×1.5 mm的分辨率采集。7〜8年后，一些受试者进行了第二次纵向扫描（[Taki等，2011](javascript:void(0);)）。第二次扫描未包含在我们的训练集中，但用于评估（请参阅结果部分）。

**OASIS**。350个随机对象。这个包含成人寿命的416个受试者的队列包含一系列每个受试者的3〜4个T1加权MPRAGE图像（尽管我们只保留第一图像），分辨率为1×1×1.5mm（[Marcus等人， 2007](javascript:void(0);)）。OASIS还发布了一个未用于此次培训的纵向队列（[Marcus et al。，2010](javascript:void(0);)）。

**ABIDE**。900个随机对象。自闭症脑成像数据交换（ABIDE 1）（[Di Martino等，2014](javascript:void(0);)）旨在研究自闭症谱系障碍，包括对fMRI静息网络的影响。这1112个主题数据集实际上是独立于美国和欧洲多个（17个）站点扫描的数据汇编。因此，T1图像在各种分辨率下包含多种对比度。就我们的目的而言，我们随机挑选了对象，对采集点和诊断不知情（ASD或不）。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.2.2。高精度单张图像

为了给我们的模型提供精确的分割知识，除了从FreeSurfer处理的实际数据中获得的训练实例之外，我们通过增强单个图像创建了更高分辨率的单个图像和相应更高精度的手动执行的海马分割目标。

我们通过平均我们自己的“MPRAGE重复测量”数据集创建了高SNR，更高分辨率（0.6 mm）的大脑图像。该数据集包括单个受试者，在各种T1-MPRAGE对比下，使用3T MRI扫描123次，改变参数如TI，TR，体素大小或并行成像量。虽然这些采集的预期目标是量化T1序列变化下的形态测量算法的行为，并与扫描持续时间进行权衡，但在这里我们使用数据集作为通过平均提供高SNR图像的方式。具体而言，对35个最高分辨率采集进行重采样和平均，以0.6 mm各向同性分辨率产生高SNR输入图像。尽可能准确地跟踪可见的海马边界地标，然后手动分割该图像。我们推断ConvNet可以认识到这些边界的有用性，并将这些知识用于其细分模型。由于FreeSurfer派生的示例集的分辨率有限，部分音量效果，来源的质量可变以及自动分割的不准确性，该知识可能无法在FreeSurfer派生的示例集中清晰地包含。

海马分割可能具有挑战性。已经描述了指南和地标（[Geuze等人，2005](javascript:void(0);)，[Malykhin等人，2007](javascript:void(0);)，[Thomas等人，2008](javascript:void(0);)）以分割单个病例。然而，由于我们的综合多平均图像具有更好的信噪比和对比度，因此客观分割它们相对容易。实际上，海马旁回的海马边界在图像中是清晰的，并且与心室角的边界也具有良好的对比，如同Pulvinar的后边界一样。在海马 - 杏仁核交界处，我们遵循视觉可观察的高信号边界。其他边界使用了alveus，fimbria和其他相对高强度。我们（BT）对ITKsnap进行了分割（[Yushkevich等人，2006](javascript:void(0);)）（<http://www.itksnap.org/>），使用三轴视图来更好地理解每个体素标记。只有这张单张图片是用手标记的。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.2.3。数据增强

我们开发了一个数据增强系统，以生成每个图像和相应目标的众多变体。该系统改变了其输入的几何形状，边界对比度和总体强度。我们将这一程序应用于队列数据（每个科目2次）和我们的高精确度科目（10,000次）。首先，对于空间变化，使用[ANT](javascript:void(0);)先进的整经机器生成随机变形场并将其应用于图像（[Tustison and Avants，2013](javascript:void(0);)）。这个变形场包含随机数（10到50之间）的随机空间散乱的各种点差的高斯失真。此过程会生成原始图像的扭曲但平滑的版本。这些扭曲中的一些故意过于极端而不切合实际，但仍然容易被人类描述，从而为我们的模型提供了相关的培训指导。其次，计算图像的拉普拉斯算子，在-2和2范围内随机加权，然后求和回图像。这改变了对比边界，从原来的锐利到几乎看不见。最后，随机生成平滑的强度偏差并将其添加到整个图像中。自然地，只有几何变换被应用于相应的目标掩模。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

2.3。履行

核心的ConvNet 通过使用CuDNN 5.1的Lasagne库在Theano（[Theano开发团队等，2016](javascript:void(0);)）上实施，并在近期（2016年末）的NVIDIA GTX 1080 GPU上进行了三周的计算培训。所有增强都是在CPU上离线进行的。预注册脚本依赖于FSL FLIRT 6.0或ANT 2.1.0。定制图像处理和统计分析使用Python软件包Scipy（[Jones et al。，2015](javascript:void(0);)），Nibabel，Pandas（[McKinney，2012](javascript:void(0);)）和Statsmodels（[Seabold and Perktold，2010](javascript:void(0);)）进行。该程序可在[https://github.com/bthyreau/hippodeep上找到](https://github.com/bthyreau/hippodeep)。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.结果

[图2](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952866/gr2.jpg)的左上[图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952866/gr2.jpg)通过我们的数据增强方案显示由同一来源产生的四种合成海马变体。该图的其余部分说明了从我们目视检查的数十个随机测试对象中挑选的各种测试对象分割。结果似乎是合理的，因为分割趋向于沿着界标，并且即使利用低分辨率信息也可以准确定位杏仁核边界。还有一些失败案例，如图左下方所示。我们的数据中发现了另一个类似的失败案例。我们也注意到一个有趣的成功：精确的海马分割，尽管大脑注射错误，大脑的最佳部分缩放到只包括海马的盒子上，产生了在训练期间从未看到的数据形状。

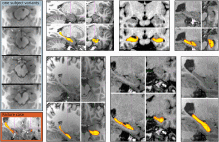
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952866/gr2.jpg)

图2

左上：数据增强的一个例子：使用我们的增强方案处理同一个源产生的四个变体。右和左下：使用我们的方法使用不同质量来源的分割示例。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952866/gr2.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952875/gr2_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr2.jpg)

对于本节剩余部分，我们进行了验证实验，并将结果与​​FreeSurfer进行了比较。我们在大量的脑部MRI图像上运行了我们的程序。每个大脑可以在约30秒内处理（大致可分为10％I / O图像读取，45％近似配准，20％Theano初始化，20％ConvNet推断和5％其他任务）。以下实验中呈现的所有结果均使用相同的最终模型进行计算，仅在未包含在训练集中的任何形式的MR图像上评估。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.1。空间范围

尽管本手稿的一个主要目标是展示从复杂软件输出中学习，但从用户角度来看，同样重要的是要显示分段任务本身的性能，而不是直接与FreeSurfer比较。由于我们无法期望获得对任何大型队列的所有科目进行的黄金标准手动分割，因此我们转向[Boccardi等人](javascript:void(0);)的杰出工作[。（2015](javascript:void(0);)），他们按照标准的协调方案对ADNI MRI队列（[Weiner et al。，2012](javascript:void(0);)）的一个相当大的子集（N = 135）进行了手动分割。我们使用MNC MR图像和在其网站上发布的nifti栅格化标签（<http://www.hippocampal-protocol.net/SOPs/index.php>）。该样本涵盖一系列扫描仪制造商（通用电气，飞利浦，西门子）和两个领域的优势（1.5T和3T）。

为了比较分割的空间范围，我们计算了一个体素DICE重叠系数。[图3](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952867/gr3.jpg)的第一行。描绘了所有135个科目的系数分布。该模式位于0.85，大多数（91％）受试者高于0.8。为了获得更多的定性理解，我们描绘了两个主题的结果。虽然似乎由于重采样误差导致一个主体（左侧，最低重叠）的手动地面实况丢失了一些体素，但我们观察到差异通常位于标签的边缘（右侧，典型范围内） 。

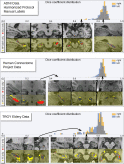
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952867/gr3.jpg)

图3

我们的方法与手动标记的ADNI（顶行）或Freesurfer标记的人类Connectome项目数据（中间行）和TRGY老年人数据（底部）的空间体素重叠。为了定性理解差异，描述了选定的案例。图中显示了分段掩码（颜色编码为：黄色=手动/自由闪存输出特有的体素，红色=对我们的方法输出唯一，棕褐色：不是任何输出的一部分）。（为了解释这个图例中的颜色引用，读者可以参考这篇文章的web版本。）

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952867/gr3.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952876/gr3_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr3.jpg)

然而，在大型队列研究中，通常依靠自动化工具。我们想比较我们的ConvNet学习方法和直接使用FreeSurfer的结果和稳健性。为了进行比较，我们使用了HCP群体的研究对象，因为该项目已经利用其强大的*Connectome管道*（[Glasser et al。，2013](javascript:void(0);)）预处理了数据。HCP数据质量高，对比一致。此外，它们的相关分段具有原始空间分辨率（0.7毫米），高于默认的FreeSurfer分辨率（1毫米），使这些数据成为空间重叠比较的良好材料。我们计算了未包含在我们训练集中的所有547个科目的得分。

[图3](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952867/gr3.jpg)（中）描绘了重叠分数的分布并且包括用于说明的相应的MRI切片。分布模式位于0.83（左0.82，右0.84）。我们的方法始终比FreeSurfer产生更小的分段。描绘了一个典型的例子（HCP 248339，右图），显示了边界的一致效果。在视觉上，人们可能会争辩说，我们的方法在这里更加精确。相比之下，HCP 205826描绘了一种情况，其中我们的方法在前区（红色区域）过度分割了左侧海马，因此具有重叠分数分布的最低值之一。一些异常值包含明确的故障，例如HCP 486759（左），FreeSurfer导致明显的错误分割，两种方法都无法分割海马尾部。

HCP数据是高分辨率质量数据，该队列中的一些数据（尽管不同科目）是我们训练集的一部分; 因此，这可能是我们算法的最佳条件。然而，我们也有兴趣在更具挑战性的条件下检查行为。敦贺市的项目（**TRGY**）由*n*  = 112 （**54F**）的日本老年人参与者组成（年龄69-75岁，平均72.15岁）。使用SPGR以0.5特斯拉获得图像，体素尺寸为1×1×3毫米。本队列研究对象中没有包括在我们的训练集中。我们使用FreeSurfer 5.3处理所有对象，将“aparc + aseg”输出标签贴图重新抽样到本地体素空间，然后像之前一样计算重叠分数。

[图3](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952867/gr3.jpg)（下）显示TRGY结果。该分布与HCP具有相同的模式，但分布更广。边缘效应在这个老年人低质量图像队列中加剧（最右边）。我们的方法有时会低估海马体的大小，因为该模型会将边界内的体素分配到阈下概率。FreeSurfer也可能发生分段错误，例如在海马面罩泄漏的左侧例子中，尽管仍有视觉可识别的边界。总的来说，尽管参与者的年龄和不寻常的海马（在我们的模型训练期间没有看到），但我们的方法通常取得了合理的结果。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.2。海马体积

接下来，我们将我们的方法的体积估计与FreeSurfer估计相比较。与重叠度量不同，总体积对实际分辨率和边界效应应该不那么敏感。此外，FreeSurfer将部分体积效应（来自*norm.mgz*）考虑到其体积计算中，这解释了为什么最终大小估计与标记体素的简单总和不同。

我们从以下来源收集海马体积数据：

**HCP**：项目提供的547个剩余科目（共计897个减去350个培训）和统计文件（aseg.stats）。

**AOBA**：317个数据点，对应245个测试科目和72个纵向数据采集。他们使用FreeSurfer 5.2横断面管道进行处理。

**TRGY**：在上一节中描述并用FreeSurfer 5.3处理的111个主题（一个失败的FreeSurfer）。

**OASIS**：58个测试主题，使用OASIS作者网站（FreeSurfer版本centos4-dev-20061005）上提供的FreeSurfer处理归档文件内的统计文件。FreeSurfer对其估计的所有3〜4T1收购进行平均，但我们仅在我们的分析中使用了第一次（001.mgz）收购。

**ABIDE**：197个测试科目，其统计文件从“ABIDE预处理”项目（<http://preprocessed-connectomes-project.org/abide/>）下载。

[图4](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952868/gr4.jpg)比较了五种数据集中两种方法的海马体积。我们观察到两种方法之间的明显相关性，以及一些异常值。当汇集数据库时，最大相关系数为0.87 / 0.82（左/右海马）。个别地，相关性是（L / R）：ABIDE 0.80 / 0.77，HCP 0.85 / 0.90，AOBA 0.82 / 0.87（或折算纵向双倍时为0.80 / 0.86），OASIS 0.95 / 0.92和TRGY 0.71 / 0.81。

[4打开大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952868/gr4.jpg)

图4

通过五个不同的数据集比较海马体积与FreeSurfer（x轴）和我们的方法（y轴）的估计值。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952868/gr4.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952877/gr4_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr4.jpg)

我们绘制了一条线性回归线。尽管这样做意味着测量误差仅来自y轴（我们的估计）这一明显不正确的假设，但我们认为定性比较不同队列仍然有用。斜率范围从0.6（ABIDE）到0.95（OASIS），TRGY介于（0.85）之间。 尽管它们的对比度不同，但最大的两组（AOBA和HCP）的斜率为0.667（无显着差异，*P* = .97）。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.3。不对称

为了更好地理解这两种方法的结果，不仅针对对方，而且针对更客观的标准，我们接下来侧重于横向不对称分析。我们推断不对称数据值得调查; 此外，两种方法都不反映的不对称性可能很好地表明潜在的分割错误。

[图5](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952869/gr5.jpg)显示了跨越群组的两种方法的右减负量。边际分布应当正态分布在0附近。当数据位于对角线上时，两种方法就参与者的不对称值达成一致。当排除异常值（通过任何方法截断为不对称最大差异的+/- 1000）时，方法之间的相关性达到*r*  = 0.55。从视觉角度来看，在非对角线数据中，FreeSurfer轴上的分布尾巴更大，这意味着它有更多的异常值。

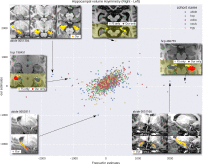
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952869/gr5.jpg)

图5

使用这两种方法分布海马体积a对称（右 - 左）。一些有趣的数据点被描绘为切片，使用与之前相同的颜色代码。没有描述范围之外的两个FreeSurfer异常值。没有数据点被图片隐藏。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952869/gr5.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952878/gr5_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr5.jpg)

为了说明，描述了具体情况。我们分割的失败在主题ABIDE 0050311清晰可见，海马头部缺失。另一个较不严重的故障案例在TRGY上部数据中显示（[图5](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952869/gr5.jpg)）。在这里，我们的方法低估了这位老年受试者的左侧体积，这可能是由于阈值过高导致的概率，可能是因为受试者与训练集不同（见讨论）。另一方面，关注主题ABIDE 0051166，视觉检查显示增加的运动伪影，这影响了右侧海马（[图5](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952869/gr5.jpg)， 右下）。我们的方法合理分段。在其他情况下，FreeSurfer由于明显的视觉错误分割低估了一侧的体积（受试者HCP 198451和HCP 486759）。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.4。在OASIS纵向测试 - 重新测试

纵向情况下通常进行海马形态测量。为了评估我们的方法在这项任务中的表现，我们依赖OASIS纵向数据库，该数据库包括150名60-96岁的受试者的数据库，其中包含多个纵向扫描。一个实际问题是“许多受试者是OASIS数据集的一部分，但被分配了新的随机标识符”（[Marcus et al。，2010](javascript:void(0);)）。为了排除其第一时间点数据是我们训练集的一部分的受试者，我们从数据本身（通过基于ANT的登记适合度评分计算图像相似性距离）估计两组之间可能的对应关系。此外，我们只保留3个或更多时间点的科目，这导致了14个可用科目。

[图6](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952870/gr6.jpg)描绘了海马体积估计与参与者年龄的关系。类似于OASIS横截面，每个扫描会话包含多个顺序采集的T1图像以提高SNR。但是，我们并没有对它们进行平均，而是提供“立即”重测信息。在大多数参与者中，年龄较大的海马缩小是清晰可见的并且是双侧观察的，这表明我们的方法可以检测到纵向变化。有时候，会期内的差异可以涵盖纵向变化（例如0007 Right），尤其是在短时间内。这可以证明MR成像过程中进行多次扫描和多个时间点的时间。

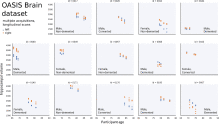
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952870/gr6.jpg)

图6

使用我们在14个OASIS测试对象上的方法评估海马体积与年龄的关系。包含多个连续收购的所有会话描述了三个或更多时间点（Marcus，[2010](javascript:void(0);)）。双方被绘制。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952870/gr6.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952879/gr6_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr6.jpg)

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.5。纵向差异

7年后，AOBA参与者中有相当数量（378人）返回进行了纵向扫描。我们使用这些数据来说明分割结果。排除所有受试者作为训练集的一部分后，我们剩下71名可用受试者（31名男性），年龄第一次扫描范围为21至70岁，平均纵向延迟时间为7.53（+/- 0.53）年。对于他们每个人，我们都运行FreeSurfer和我们的方法，两个时间点被独立处理。

在[图7中](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952871/gr7.jpg)，左图显示估计的纵向海马体积（为了清晰起见，仅绘制左侧）。使用这两种方法可观察到海马体衰老的总趋势，这在男性中更为显着。使用我们的方法，FreeSurfer细分（47岁的女性受试者，不可信的海马减少）没有出现。视觉检查揭示了将海马的一部分错误分类为皮质组织。在另一个科目的海马右侧发生了类似的问题（未显示）。图中的右图明确比较了海马纵向变化的估计值（即第二个减去第一个时间点）。一个理想的协议将使所有数据都在对角线上。排除上述两个异常值，相关性为r = 0.70（左）和r = 0。61（右图，未显示）。由于FreeSurfer的海马体积绝对估计值一般比我们的方法高，因此我们预计纵向估计值将略低于对角线上的平均值，这在图上可以观察到。最后，我们注意到FreeSurfer具有纵向流水线; 尽管我们没有使用它，因为它的计算需求很大，因为它不在我们验证实验的范围之内，它可能在两个时间点之间提供了正常化的体积估计。

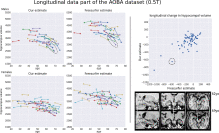
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952871/gr7.jpg)

图7

左侧面板：海马体积（使用两种方法只显示左侧）。上排：31名男性，下排：41名女性; 左：我们的方法，右：FreeSurfer方法。仅为了清晰而添加颜色。右侧：海马体积纵向变化（平均7.53 + / - 0.53年后）。示例主题（右下）被描绘为与其他图中用虚线圈出的数据点相对应的共同登记的切片。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952871/gr7.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952880/gr7_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr7.jpg)

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

3.6。对比度容差

对于最后的测试，我们想了解我们的ConvNet识别对比度变化的稳健性。我们决定使用整个重复MPRAGE数据集。如前所述，该数据集包含在各种序列设置下扫描123次的单个主题。然而，我们强调这个主题的一些收购被用来生成训练实例，因此ConvNet已经熟悉这个主题的形态，所以结果会有偏差。然而，我们决定继续进行，因为我们对序列变异的鲁棒性特别感兴趣，并且这些变化出现在训练时间。

[图8](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952872/gr8.jpg)总结了结果。训练集的大对比度变化似乎提供了一个有用的基准，这确实表明了我们算法的行为变化。理想情况下，该图应该有两个位于真实L / R体积值的单个点。该图显示我们方法的体积变化范围小于FreeSurfer估计的范围。（另见补充材料B）此外，我们的方法在左右差别非常一致（配对t检验t = 23.42，P = 0），而FreeSurfer不是（配对t检验t = -0.83，P = 0.41）。一些具有失败分割的异常值对应于非常具有挑战性的测试图像（例如，在小于30秒内通过高度平行并行成像获取的MPRAGE），如下面的示例切片所示。

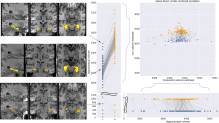
[](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952872/gr8.jpg)

图8

估计海马体积变化的结果与对比度变化，但没有解剖变化。理想的结果应该包含两个（L / R）单个点的真实体积值。结果可能偏向于我们的方法（见正文）。

[查看大图](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952872/gr8.jpg) | [查看高分辨率图像](http://www.medicalimageanalysisjournal.com/cms/attachment/2119159546/2089952881/gr8_lrg.jpg) | [下载PowerPoint幻灯片](http://www.medicalimageanalysisjournal.com/action/downloadFigures?pii=S1361841517301597&id=gr8.jpg)

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

4。讨论

我们提出了一种从人脑T1图像中分割海马的方法，在机器视觉社区最近取得成功之后，以卷积神经网络为核心。我们的神经网络使用监督式学习进行训练，部分来自从多个大型私人和公共数据集获得的实际数据。至关重要的是，我们使用现有算法（FreeSurfer）的输出作为监督，而不是黄金标准的手动标记输入。这使我们能够获得比其他方式更大的训练集。训练数据的另一部分来自模拟数据，它们以高精度训练网络，虽然是合成的特征。我们使用基于随机生成的非线性平滑翘曲场和模拟对比度和强度效应的新型数据增强方案。最后，

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

4.1。软件适用性

过去的十年中，许多大型队列脑扫描项目已经到来。另外，通过联合来自现有较小群组或医院组的多个现有数据集来创建新的元数据集。这反过来又创造了对广泛适用的工具的需求，对群组变异的强大性，以及资源的高效性。

但是，目前大部分围绕机器学习发表的大脑图像分析工作都依赖于小而精确的图像集。在这篇论文中，我们依靠大量数据集的存在，这些数据集具有固有的大的MR对比度或人口统计学差异，以创建一个稳健的模型。该模型直接适用于新的队列，无需任何重新培训。考虑到重新培训和重新验证ConvNet模型所需的时间，专业知识和数据，这一点很重要。该模型的当前版本及其受过训练的参数可以在线获得，其形式为直接可用的软件程序，无需特殊调整，也无需外部地图集。

此外，该方法对于以更快的方式分割海马实际上是有用的。ConvNets推理速度非常快，即使在较早的GPU硬件上也是如此。目前的海马分割工具相对较慢; 例如，FSL FIRST需要10分钟以上，FreeSurfer需要4小时以上（对于统计量计算）。虽然后者不会将其分析限制在皮质下结构上，但许多研究集中于海马，这是队列研究中人脑最广泛研究的亚结构之一（例如[Janowitz等，2014](javascript:void(0);)）。在很多情况下，等待几个月的计算来处理几千个主题是不切实际的。虽然存在其他替代方法，但我们认为我们的卷积神经网络的快速推断可以有效地对有效的海马分析作出贡献。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

4.2。模型设计

该模型的核心是卷积神经网络。我们通过反复试验来调整宽度和深度，并根据时间和硬件进行权衡。虽然与其他领域中使用的网络（如自然图像处理）相比较相对较浅，但其大小恰好与其他医学图像处理任务中使用的网络相当。值得注意的是，我们注意到添加更多图层似乎并没有改善结果。然而，由于我们没有足够的资源来全面探索每个设计决策的影响，所以我们依赖最佳实践，如可用的最佳实践，例如剩余块中的层次顺序或3的内核大小。

我们方法中的一个重点是我们增强数据的设计。对于每个真实的大脑数据，我们使用多个弹性变形生成几何失真的实际数据版本，并通过模拟对比度变化改变外观。我们故意将这些变化调整到超出医学上可信值的范围之外，因为执行的任务 - 结构的分割 - 在这些极端情况下仍然很好定义。这在概念上类似于看夸张的例子实际上可以帮助理解问题。这些改变使得网络相当强大，可以做出大量的变化，包括在训练集中从未见过的变化。

尽管如此，我们的训练集可以改进以包含更多种类。事实上，我们现在认为，我们花费的时间超过了必要的训练网络的时间，以反对许多彼此过于相似的人，尽管多次小规模实验旨在解决培训期间增强，模拟和对比度变化的影响证实了大数据集的有用性（见补充材料A）。虽然我们的数据增强程序对于产生一系列可能的和极端的几何变异性很有用，但它不能专门模拟一些现有的解剖变异性，例如在医疗条件下观察到的皮质收缩或心室空间扩大。在未来的版本中，我们将更加重视将异常的大脑包含在训练集中，

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

4.3。验证

为了进行验证，我们计算了一个针对手动分割的ADNI子集的重叠度量，然后我们将该方法应用于超过一千个未在训练时看到的图像。由于海马分割方法的失败率较低，因此这是必要的。我们将注意力集中在与空间范围，不对称性，纵向变化和序列参数容差有关的分布的异常值上。

由于大型队列没有可用的实际情况，并且由于难以对数千个图像进行视觉检查，因此我们在我们的程序和常用的FreeSurfer之间进行了广泛的度量比较。结果在个体水平上是非常合理的，并且在大群体中与FreeSurfer有很好的相关性，表明我们的程序可以同样用于群体建模。通过查看异常值，我们可以确定FreeSurfer失败的许多角落案例，以及少数我们的方法失败的案例。这很有趣，因为据报道FreeSurfer（和FSL-First）显示出与手动分割（[Mulder等，2014](javascript:void(0);)）相同的重现性，至少用于监测纵向变化。

在运行ConvNet分割之前，我们的方法依赖于近似注册步骤。由于技术上的限制，这些步骤是非常必要的：将输入数据的大小减小到更小的区域，以使其适合内存。未来的版本可能会放宽这个约束，或者可能以另一种方式达到这一点，包括使用ConvNet（[Miao et al。，2016](javascript:void(0);)）进行本地化。在当前版本中，我们重新使用FSL FLIRT或ANT。虽然这两种算法提供了高质量的注册，并且依赖于许多竞争性分割方法（例如[Pipitone等，2014](javascript:void(0);)），这里我们只将它们用于初始的低分辨率仿射映射。这意味着我们的方法最终对MRI注册中不可避免发生的中小误差不敏感。

尽管如此，在其余一些情况下，可能会发生重大错误，例如无法正确对齐轴。由于我们希望我们的方法在大群体中稳健，我们通过将管道加倍，使用FSL FLIRT和ANT作为替代（FSL有4％的失败率，然后用ANT恢复了所有两个数据），从而减轻了这种影响。比做相反的效率更高）。这是可能的，因为主要的注册错误很容易被自动检测。尽管我们最初专注于检测这种失败转换的异常配准矩阵雅可比行列式，但后来我们注意到，仅仅检查ConvNet的输出就暴露了更明显的错误特征。此外，我们相信使ConvNet能够在未来的版本中直接输出信心得分很容易。

我们也通过检查异常值来欣赏我们方法可能的失败。排除完全未能注册的两个案例，我们仍然能够发现一些分割不当的案例。在两种情况下，海马头与杏仁核混淆，可能是由于额外的超强度，或窄边界框。在另一种情况下，一个老年受试者，海马体素被分配了低于阈值的概率，可能是由于形态与训练图像差别太大，再加上倾斜的后注册边界框。最后，在其他情况下，数据中会（特意）出现极大量的噪音，这会混淆ConvNet。事实上，对深层神经网络的一次反复批评是，他们如何以及为何产生结果缺乏透明度。努力输出可解释的结果确实存在; 然而，可以说这是为了达到超越简单模型所能做到的准确性而付出的代价。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

4.4。学习和转移医疗成像

过去十年来，许多大型队列神经影像项目的到来，最有趣的是，这些数据的公开可用性。这些最近的共享计划（[Poline等，2012](javascript:void(0);)）为基于机器学习的方法提供了大量机会。

在这份手稿中，我们演示了从复杂软件包中监督学习。与从手动标记的基本事实中学习相比，这种方法具有能够训练更广泛的数据集的明显优势。这直接关系到系统的最终健壮性。目前的机器学习方法仍然需要在达到可接受的性能之前看到大量的样本。只要他们不能从几个例子中复制令人印象深刻的人类泛化技能，另一种方法就是依靠大群体的丰富性。

与直接运行FreeSurfer本身相比，我们的方法速度更快。但通过查看测试集中的异常值，它也可以产生更少的失败案例。值得注意的是，我们的模型在FreeSurfer输出上进行了训练，在应用新数据时没有产生类似于FreeSurfer算法的错误。当然，ConvNet从算法输出中学习，而不是从算法本身。但这仍然令人惊讶，因为一些类似的FreeSurfer失败案例不可避免地出现在训练集中。最近有研究表明，一个参数比训练目标更多的ConvNet可以学习任何映射，有意义或无意义（[Zhang et al。，2016](javascript:void(0);)）并且仍然可以推广到可能的程度。我们直观地证实，在训练集中，一些FreeSurfer错误分割似乎已被记忆。在测试中运行新数据时不会发生类似的错误。这可能是因为嵌入在网络中的知识是从广泛的数据源中获得的，并且由于输出是复杂的（分段，而不是分类）和输出的网络容量是有限的，因此在很大程度上正规化了。这证明了ConvNets有效编码可变置信水平的多个来源的能力。此外，尽管目前还不清楚为什么受过训练的ConvNets概括得如此之好，[Zhang等人](javascript:void(0);)的工作[。（2016）](javascript:void(0);)表明它们可能形成一个高效的分层局部模式匹配系统。在这种推测性解释之后，我们可能会将我们的系统看作是多图集匹配的极端形式。

最后，除了上面提到的直接利益之外，我们认为存在一些进一步的优势。在这篇手稿中，我们主要从一个单一的软件（FreeSurfer）中学习，但它可以以与来自多个软件源的数据完全相同的方式应用，结果可能比任何单独的软件更好，因为合奏方法经常显示。更根本的是，通过学习软件包，我们将手工制作的算法知识转换为更通用的形式：嵌入到神经网络中。神经网络可以很容易地在更大的环境中插入和重用，以执行新的任务，可能有一些以任务为中心的微调和关注机制。因此，这里提出的方法可以被看作是朝着来自不同软件源的知识的轻松混合迈出的一步，

我们在这里介绍的方法与海马分割问题相关。我们的研究表明，深度神经网络可以从更复杂的手工算法中学习部分特征。ConvNets可以从一系列具有潜在缺陷的样本中提取有意义的知识。我们相信，我们在此采用的知识提取和建模方案可以应用于许多医学问题。

跳转到部分1.介绍2.材料和方法  2.1。分割模型    2.1.1。预处理    2.1.2。损失函数和优化参数  2.2。培训数据    2.2.1。队列图像    2.2.2。高精度单张图像    2.2.3。数据增强  2.3。履行3.结果  3.1。空间范围  3.2。海马体积  3.3。不对称  3.4。在OASIS纵向测试 - 重新测试  3.5。纵向差异  3.6。对比度容差4。讨论  4.1。软件适用性  4.2。模型设计  4.3。验证  4.4。学习和转移医疗成像致谢附录。补充材料参考

致谢

AOBA数据集由日本电信发展组织，国家精神卫生研究所，国家神经疾病和中风研究所，国家药物滥用研究所和美国国家癌症研究所提供资助。我们还受益于外部大型共享处理数据集的存在：OASIS数据集受益于以下授权号：P50 AG05681，P01 AG03991，R01 AG021910，P20 MH071616，U24 RR021382。人类连接组计划（WU-Minn Consortium）（首席研究员：David Van Essen和Kamil Ugurbil; 1U54MH091657）由16个NIH研究所和中心资助，支持NIH神经科学研究蓝图; 并由华盛顿大学麦克唐奈系统神经科学中心提供。至于ABIDE I：“NIMH（K23MH087770）和Leon Levy基金会为Adriana Di Martino的工作提供了主要支持。Michael P. Milham和INDI团队对该项工作的主要支持是由Joseph P. Healy和Stavros Niarchos基金会向儿童智力研究所提供的赠品以及向MPM提交的NIMH奖（R03MH096321）提供的。