3 d U2-Net:用于多领域医学图像分割的三维通用U-Net?

黄超1,2[0000−0003−4974−2649]，韩虎2,3，姚青松2，朱善宽1()，周s . Kevin 2,3()

1 浙江大学医学院公共卫生学院、妇女医院慢性病研究所，营养与食品卫生系，浙江杭州310058

zsk} {huangchao09, @zju.edu.cn

2 医学影像，机器人技术，分析计算实验室/工程(奇迹)，中国科学院智能信息处理重点实验室，中国科学院计算技术研究所，北京100190

3 深圳鹏程实验室

(套arXiv: 1909.06012 v1。IV] 2019年9月4日

摘要像U-Net这样的全卷积神经网络已经成为医学图像分割的最先进方法。实际上，一个网络是高度专门化的，并为每一个分割任务单独训练。与其学习多个模型的集合，不如学习不同任务的通用数据表示，理想情况下是学习一个单一的模型，为每个任务添加最少数量的参数。受最近多领域学习在图像分类中的成功启发，我们首次探索了一种有前途的通用架构，该架构处理多种医学分割任务，并可扩展到新的任务，无论不同的器官和成像方式。我们的3D Universal U- net (3D U2- net)建立在可分离卷积的基础上，假设来自不同领域的图像具有特定于领域的空间相关性，可以通过通道卷积来探究，同时也共享跨通道相关性，可以通过点卷积来建模。我们计算3D U2-Net在五个器官分割数据集上。实验结果表明，该通用网络在分割精度方面能够与传统模型竞争，而所需参数仅为1%左右。此外，我们观察到体系结构可以轻松有效地适应新的领域，而不会牺牲用于学习通用网络共享参数化的领域的性能。我们把代码3D U2-Net进入公共领域。4

关键词:通用模型多领域学习分割

4

吗?

Huang C.和S. Zhu得到了Cyrus Tang基金会和浙江大学教育基金会的支持。国家自然科学基金项目(61732004和61672496);中国科学院对外合作项目(GJHZ1843);中国科学院青年创新促进会项目(2018135)。这项工作是C. Huang在MIRACLE实习时完成的。

https://github.com/huangmozhilv/u2net\_torch/

黄等。

2

1介绍

图像分割是临床实践和健康研究的关键。U-Net[15]等全卷积神经网络(CNNs)已经成为医学图像自动分割的主导方法[4,11]。一个实用的分割模型是通过为某个任务或数据集定制一个神经网络体系结构，然后从无到有地进行训练来学习的[11,16,18]。[7]为不同扫描仪和/或协议获取的大脑数据集学习了单个分割CNN。尽管这些模型功能强大，但由于高度专业化的设计，这些模型很难扩展到包含看不见内容的新任务。[6]更进一步，为各种任务提供了一个自适应框架，为每个任务生成了相互独立的模型。相反，人类专家可以很容易地学会处理多重任务，并在获得技能的基础上归纳出新的任务。之前有多篇文章探讨了多任务分割，即所有感兴趣的器官都出现在同一幅图像中[9,17]。这里我们考虑一个更现实和更具挑战性的场景:对于给定的数据集，只扫描人体的局部区域，并且图像中只有一个或几个解剖结构被注释。[12]专注于一个类似的主题，在三个任务上训练一个CNN，但是训练的模型被设计成不能扩展到其他任务。从这个角度来看，如何有效有效地进行图像分割仍然是一个有待解决的问题。

Bilen等人[2,13,14]认为可能存在跨不同视觉域的通用数据表示。具体来说，他们引入了一种新的竞赛，叫做视觉十项全能挑战赛 5，旨在同时建模10个不同风格和内容的视觉域，如互联网图像、手写文字、素描、浮游生物等。他们提出了“多领域学习”这一新课题，并通过在Ima- geNet预先训练的模型上搭载并行剩余适配器来实现通用表示。然而，他们的工作只专注于图像分类。自然，一个问题出现在我们面前:是否有可能建立一个单一的神经网络来处理不同领域的医学分割任务?

为了实现这一目标，我们从以前的研究中获得了灵感[3,5]，特别是迄今为止获得视觉十项全能挑战第一名的[5]。[5]认为[14]忽略了各个领域的结构异质性，试图利用深度可分卷积来解决这个问题。标准卷积同时进行空间和通道的计算，这种卷积将计算分为两个连续的步骤:首先，深度卷积对每个输入通道应用一个独立的卷积滤波器，然后点卷积对每个空间位置的所有通道的输出进行线性组合。他们的多域网络的基本构建块包括一个并行的信道卷积队列，每个域一个，然后是一个所有域共享的点卷积。前者更好地捕捉特定领域的空间模式，而后者探测可共享的跨通道相互依赖-

5

https://www.robots.ox.ac.uk/~vgg/decathlon/

3 d U2-Net:用于多域分割的三维通用U-Net

3.

dencies。在本文中，我们声称将“纵深可分离卷积”标注为“可分离卷积”，“纵深卷积”标注为“通道卷积”，以避免与图像体积的深度维度混淆。

在前面介绍的可分离卷积的基础上，我们的工作提出了一种通用的多域医学图像分割体系结构。其主要思想相当直观而强大:首先在3D U-Net[4,15](或V-Net[11])的基础上设计一个基础网络，然后将任何步幅为1的3 × 3 × 3标准卷积替换为类似[5]的可分离卷积。然而，我们的方法与[5]有很大的不同:(1)他们的工作侧重于图像分类，这与我们这里的图像分割有本质的不同。(2)通过三个步骤得到最终的多域体系结构:首先，在ImageNet上预训练一个经过可分卷积的ResNet-26;第二，冻结并将点向卷积权值转移到新的网络中;第三，在每个域上分别训练新网络，并将通道卷积叠加在一起，同时共享预先训练模型的点卷积权值。然而，我们设法跨领域一起训练，以获得最终的模型。(3)通过简单地添加新的信道卷积，我们进一步使我们的通用网络适应新的领域。据我们所知，这是第一次学习一种用于多领域医学图像分割的可扩展通用网络。

2方法

2.1问题定义

让{D1D2···，DT }为T个图像域的集合，其中Dt 由两个成对的图像空间{XtYt}。Xt ∈RCt×D×H×W是输入



图像空间和Yt ∈RCt×D×H×W为输出图像空间，即:,分割

面具。D、H、W分别为空间深度、高度、宽度。Ct 和Ct0 是特定于每个领域的成像方式和分割类别的数量。为了在所有领域都能很好地工作，我们的通用网络包含领域特定参数和共享参数。让θt 为域D的域特定参数t 和θu 是所有域共享的参数。假设{xt,i, yt,i}为定义域Dt的第i个训练对，则神经网络F(X)的输出Yˆ为



(1)

2.2域适配器

域适配器是确保我们的通用网络成功的关键组件，它由领域特定参数和共享参数组成，并建立在可分离卷积(而不是标准卷积)之上。

在滤波器W∈的标准卷积中 R3×3×3××C 0将输入张量U∈RC×D×H×W应用，得到输出张量Uˆ∈RC0×D×H×W

黄等。

4

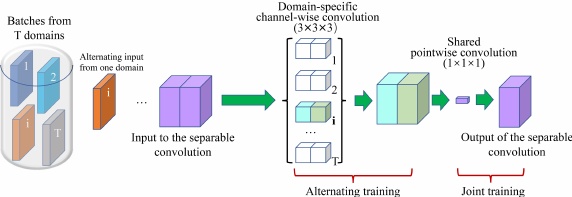
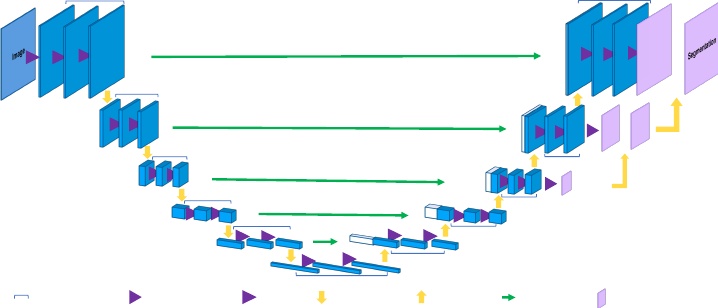


图1所示。基于可分卷积的域适配器。



一个

一个

一个一个

+

16

一个32

一个

一个一个

+

一个64

一个

一个

一个

一个

一个

128

一个

一个

一个

一个

一个

256

一个

剩余的连接

1 1 1卷积

一个

一个

512

适配器

下采样

了抽样

连接

多层次分割

图2所示。提议的3D通用U网(3D U2净)。

对输入端应用C0滤波器w∈R3×3×3×C并行连接C0输出特征映射。一个简单的计算告诉我们，上述过滤器中过滤器参数的总数是27∗C∗C0。此外，当训练模型分别为T域，参数的数量增长T倍!

在可分卷积中，计算被分解成两个连续的步骤。第一步应用C通道滤波器w∈R3×3×3 对每个通道的输入并行并连接C输出特征映射在一起。在这里，每个域都有自己的信道滤波器。第二步应用C0 逐点滤波器w∈R1×1×1×C输出C0通道的最终特征映射。这里，所有域都共享相同的逐点过滤器。一个简单的计算告诉我们，上述过滤器中权重的总数是27∗C∗T + C∗C0。图1说明了如何组装特定于域的信道卷积和共享的点卷积来形成域适配器。

2.3 3D Universal U-Net (3D U2-Net)

如图2所示，我们的通用网络体系结构是基于一个基本网络，包含六个部分:(1)输入;(2)编码器路径;(3)瓶颈块;(4)译码器路径;(5)深度监管分支机构;(6)输出。渠道的

3 d U2-Net:用于多域分割的三维通用U-Net

5

输入和输出可以根据成像方式的数目和不同领域的类别而有所不同。通常，输入层使用16个过滤器。编码器和解码器路径都包含不同分辨率的五个级别。剩余连接应用于每个级别。跳过连接用于从解码器路径[15]的编码器对应端保存更多上下文信息。受[8]启发，我们通过多层次分割映射的元素和，在解码器路径的末端加入了一个深度监督分支，以提高最终定位性能。为了构建通用网络，在基础网络中插入上述的域适配器，以1的步长代替任何标准的3 × 3 × 3卷积。

2.4损失函数

采用混合损失函数，将改善交集-union分割分数的Lovsz-Softmax损失[1]与焦点损失[10]相结合，以缓解类别失衡。在训练通用模型的过程中，我们以轮循的方式从每个数据集中抽取一批样本，允许每个域贡献共享参数。假设对于第n次迭代，批处理数据对{xtyt}来自域Dt，相应的损失Ln 是



(2)

在θt 为域D的域特定参数t 和θu 是神经网络的普遍共享参数。lL 洛夫斯-软max输了吗f 是焦损失的对应物。

3实验结果

在本节中，我们提出了广泛的实验来评估建议的3D U2(1)针对每个基域分别训练基网络，获得独立的模型，目的是再现传统方法;(2)共享模型(shared model)，研究模型的所有参数是否可以被所有域共享，通过将所有基域一起训练单个基网络来获得;(3)通用模型，这是我们的最终目标，是通过同时训练所有基本域的通用体系结构来实现的。值得注意的是，前两个代表了两种极端的多器官分割方法，并作为通用模型的基线。此外，我们还测试了共享模型和通用模型在一个新领域上的通用性。

数据集:我们使用来自医学分割十项全能挑战的六个公共数据集6 正如[19]所介绍的。前5个数据集作为基域，用于训练通用模型。另一方面，最后一个

6

https://decathlon.grand-challenge.org/

黄等。

6

表1。数据集的基本特征。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 任务 | 形态 | 数据大小图像形状 | | 立体像素间距 |
| Base01心 | 核磁共振成像 | 20. | (90∼130)×320×320 | 1.37×1.25×1.25 |
| Base02肝 | CT | 131 | (74∼987)×512×512 | (0.7∼5)×(0.557∼1)×(0.557∼1) |
| Base03海马 | 核磁共振成像 | 260 | (24日∼47)×(40∼59)×(31∼43) | 1×1×1 |
| Base04前列腺 | T2, ADC | 32 | (11日∼24)×(256∼384)×(256∼384) | (3∼4)×(0.6∼0.75)×(0.6∼0.75) |
| Base05胰腺 | CT | 281 | (37∼751×512×512 | (0.7∼7.5)×(0.605∼0.977)×(0.605∼0.977) |
| 新脾 | CT | 41 | (31∼168×512×512 | (1.25∼7.5)×(0.535∼0.977)×(0.535∼0.977) |

将数据集作为一个新的领域，用于测试通用模型的适应性。表1显示了数据集的基本特征。对于每个数据集，随机抽取80%的样本进行训练，其余20%作为测试数据。

预处理:数据集在形态、图像大小和体素间距方面高度不同。预处理过程如下:(1)将所有图像裁剪到非零值区域，从而减小图像尺寸，减轻计算负担;(2)将所有图像重新采样到对应数据集的中位体素间距，保留空间语义;(3)对于每个患者，将图像剪切到整个图像强度值的[2.0,98.0]百分位数，然后用每种模态图像的均值和标准差进行Z-score归一化;(4)应用了以下数据增强:随机弹性变形、随机旋转、随机缩放和随机镜像。数据增强是在训练期间用批生成器“实时”完成的，批生成器是由德国癌症研究中心医学图像计算部门维护的python包。

为了适应有限的GPU内存，我们从整个图像中随机抽取patches来训练网络。而对于推理来说，这些补丁是通过一个滑动窗口在整个图像上移动以半个补丁大小的步长来生成的。对于共享模型和通用模型，输入批处理为两个patches，大小为128 × 128 × 128，下采样次数设为6次。而对于独立模型，我们根据图像大小调整输入的patch大小和每个域的分辨率，以最大限度地利用计算资源。如果中值形状小于128×128×128，我们在输入的patch大小和batch大小之间切换，使patch大小与中值形状具有相同的长宽比。设置每个轴的下采样操作次数，直到最深层的feature map大小减小到8。具体来说，为了准备用于共享模型和通用模型的patch，我们首先提取一个与独立模型相同大小的patch，然后将其调整为上述目标patch的大小。

实现细节:在NVIDIA V100 GPU上使用Pytorch 1.0.1实现网络。应用ADAM优化器，初始学习速率为3 × 10−4 重量衰减为10−5。epoch定义为超过250个批次的迭代。指数移动平均，ltMA，是监测训练损失为每30个时期。只要ltMA不降低5 × 10，学习率就会降低5倍−4。一旦学习结束，我们就终止培训

7

https://github.com/MIC-DKFZ/batchgenerators/

3 d U2-Net:用于多域分割的三维通用U-Net

7

表2。碱基域的定量结果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Base01 | Base02 | Base03 | | Base04 | | Base05 |  |
|  | 心 | 肝 | 海马体 | | 前列腺癌 | | 胰腺 |  |
| (骰子%) | 左心房 | 肝 | 前后 | | PZ | TZ胰腺的意思 | | |
|  |  |  |  |  |  |  |  |  |
| 独立的 | 93.26 | 95.02 | 89.62 | 87.74 | 58.39 - 87.18 | | 78.78 | 84.28 |
| 共享 | 92.73 | 93.40 | 89.25 | 87.30 | 68.38 - 89.30 | | 57.57 | 82.56 |
| 通用 | 91.98 | 93.54 | 89.34 | 87.05 | 68.50 - 89.21 | | 62.08 | 83.10 |

利率低于10−8。在训练共享模型和通用模型的过程中，我们采用循环的方式，轮流向每个域的网络批样本输入，从而使所有域对最终模型的贡献相等。在试验数据上给出了结果。

基域定量结果:表2列出了三种模型在每个基域上的平均骰子得分。沿着列进行比较，我们观察到独立模型在大多数领域获得最高分数，并产生最高的总体平均分数。然而，令人惊讶的是，共享模型和通用模型在大多数领域都取得了与独立模型相当的中等性能，并在Base04前列腺的外周区(PZ)和过渡区(TZ)方面取得了显著的增长。与共享模型相比，我们进一步观察到通用模型在Base05胰腺的分割中表现得更好。此外，与共享模型相比，通用模型在所有领域的总体平均分数更高。总体性能的提高可以归因于使用了适用于各个领域的特定参数。

模型复杂性:在研究模型的复杂性时，我们排除了输入层、最后一层和深度监管分支，因为它们从来不跨域共享。以共享模型中使用的基本网络为参考。表3(a)计算并显示了参数的数量。显然，提议的3D U2-Net需要的参数最少，表明它可以有效地跨各种域执行。通用模型的参数总数约为所有独立模型的1%左右，但两者的分割精度相当。

定量结果的一个新领域:此外,我们进行实验来说明适应训练共享模型的有效性或通用模型到一个新的任务,由冻结相应的共享实现逐点的旋转或标准隆起和添加和培训所有其他特定于域的模块如输入层,channel-wise沟回在平行于同一家族的结构域。表3(b)显示，与共享模型相比，通用模型对新域“新脾脏”的表现更好，因此表明共享模型具有优于共享模型的泛化能力。这进一步证明了特定于领域的参数的有效性。该通用模型对具有少量额外参数的新领域具有较好的适应性。， 0.3%，这正是我们在本文中所预期的。

黄等。

8

表3。(一)模型的复杂性。(b)新脾结构域的定量结果。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (a) #Par (a)比率(b)新脾-骰子% (b) #添加Par | | | |
| 独立的126.7 | | 4.1× | 92.37 | 30.7米 |
| 共享 | 30.7米 | 1× | 90.67 | 0 |
| 通用 | 1.7米 | 0.06× | 91.60 | 0.1米 |

4结论

综上所述，我们提出了一种新型的通用神经网络3D U2基于-Net的多器官分割问题，填补了可扩展多领域学习在图像分割中的空白。实验结果表明，该方法只使用了很小一部分参数，分割性能与传统方法训练的独立模型相当。由于CT和MRI图像是现有的常规图像，而且人体器官的数量是恒定的，因此用于多器官分割的通用模型在不久的将来就可以得到充分的发展。此外，所提出的框架可扩展到许多其他多领域的应用，从而促进神经网络的临床实践。

参考文献

1.Berman, M.， Rannen Triki, A.， Blaschko, M. b .: lov´asz-softmax损失:对于神经网络中的交集-合并度量的优化而言，这是一个易于处理的替代。: Proc。CVPR。第4421 - 4413页(2018)

2.Bilen H.， Vedaldi A.。字体普遍表征:脸、文字、浮游生物和猫的品种之间缺失的联系。arXiv: 1701.07275 (2017)

3.Chollet F。使用深度可分离卷积的深度学习。: Proc。CVPR。第1258 - 1251页(2017)

4.C¸¸cek, O。，“Abdulkadir, A.， Lienkamp, s.s.， Brox, T.， Ronneberger, O.”: 3D U-net:从稀疏标注学习密集体分割。: Proc。MICCAI。424 - 432页。施普林格(2016)

5.郭勇，李勇，Feris, R.，王磊，Rosing, T.。深度卷积是你学习多个视觉域所需要的全部。arXiv: 1902.00927 (2019)

6.Isensee, J.， Petersen, J.， Klein, A.， Zimmerer, D.， Jaeger, P.F.， Kohl, S.， Wasserthal, J.， Koehler, G.， Norajitra, T.， Wirkert, S.等:nnU-Net:基于u-net的医学图像分割自适应框架。arXiv: 1809.10486 (2018)

7.Karani, N.， Chaitanya, K.， Baumgartner, C.， Konukoglu, E.。:终生学习的方法，通过扫描仪和协议的大脑MR分割。在:Proc. MIC- CAI。476 - 484页。施普林格(2018)

8.Kayalibay, B.， Jensen, G.， van der Smagt, P.。:基于cnn的医学影像数据分割。arXiv: 1701.03056 (2017)

9.Lay, N.， Birkbeck, N.， Zhang, J.， Zhou, s.k:利用上下文整合和判别模型快速进行多器官分割。: Proc。MICCAI。450 - 462页。施普林格(2013)

10.Lin, T.Y.， Goyal, P.， Girshick, R.， He, K.， Doll´ar, P.。:用于高密度物体检测的焦损失。: Proc。ICCV。第2988 - 2980页(2017)

3 d U2-Net:用于多域分割的三维通用U-Net

9

11.Milletari, F.， Navab, N.， Ahmadi, S.A.: V-Net:用于容积医学图像分割的全卷积神经网络。: Proc。3 dv。第571 - 565页(2016)

12.Moeskops, P.， Wolterink, j.m.， van der Velden, b.h.， Gilhuijs, K.G, Leiner, T.， Viergever, m.a.， Iˇsgum, I。:深度学习在多模式下的多任务医学图像分割。: Proc。MICCAI。478 - 486页。施普林格(2016)

13.Rebuffi, s.a.， Bilen, H.， Vedaldi, A.。:使用残余适配器学习多个视觉域。: Proc。冻伤。第516 - 506页(2017)

14.Rebuffi, s.a.， Bilen, H.， Vedaldi, A.。:多域深度神经网络的高效参数化。: Proc。CVPR。第8127 - 8119页(2018)

15.Ronneberger, O.， Fischer, P.， Brox, T.。U-net:用于生物医学图像分割的卷积网络。: Proc。MICCAI。234 - 241页。施普林格(2015)

16.罗斯,人力资源,L,法拉克,A。,Shin H.C,刘,J。,Turkbey,后代e.b.,萨默斯,智慧化: DeepOrgan:用于胰腺自动分割的多层次深度卷积网络。: Proc。MICCAI。556 - 564页。施普林格(2015)

17.Roth, H. r .， Oda, H.， Hayashi, Y.， Oda, M.，清水，N.，藤原，M.，三泽，K.， Mori, K.。:用于多器官分割的分层三维全卷积网络。arXiv: 1704.06382 (2017)

18.Savioli, N.， Montana, G.， Lamata, P.。V-FCNN:用于心房自动分割的容量全卷积神经网络。arXiv: 1808.01944 (2018)

19.Simpson, a.l.， Antonelli, M.， Bakas, S.等:用于开发和评价分割算法的大型注释医学图像数据集。arXiv: 1902.09063 (2019)