

基于全卷积 DenseNet 的前列腺 MRI 分割新方法*

胡学刚, 杨洪光[†]

(重庆邮电大学 通信与信息工程学院 重庆市信号与信息处理重点实验室, 重庆 400065)

摘要: 前列腺磁共振图像(MRI)的自动分割对前列腺疾病的诊断至关重要,但是前列腺区域所占比例过小、组织边界模糊等问题为自动分割带来极大困难。针对这些问题,提出了一种基于全卷积 DenseNet 的前列腺 MRI 图像分割方法。该方法以现流行的深度学习理论为基础,利用迁移学习的思想,将 DenseNet 从自然图像迁移到前列腺数据集;采用反卷积和类似 U-Net 的全卷积神经网络结构,实现端到端的图像分割。同时引入并改进 Dice 相似性损失函数以解决前列腺 MRI 中背景所占比例远远大于前列腺区域和一些像素难以准确分割等问题。通过在 PROMISE12 数据集上进行实验,提出的方法 Dice 相似性系数达到 93.25%, Hausdorff 距离小于 1.2 mm,相较于目前的主要方法,分割效果更好、所耗时间更短。

关键词: 前列腺 MRI 分割; DenseNet; 全卷积神经网络; Dice 损失函数

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1001-3695(2020)06-059-1886-04

doi:10.19734/j.issn.1001-3695.2018.11.0909

Novel prostate MRI segmentation method based on full convolution DenseNet

Hu Xuegang, Yang Hongguang[†]

(Chongqing Key Laboratory of Signal & Information Processing, College of Information & Communication Engineering, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)

Abstract: Automatic segmentation of prostate MRI images is very important for the diagnosis of prostate diseases. However, due to the small proportion of prostate regions and fuzzy tissue boundaries, automatic segmentation has become an important challenge. This paper proposed a novel prostate MRI segmentation method based on full convolution DenseNet. Based on the popular deep learning theory, the method applied the idea of transfer learning to load DenseNet from natural images to prostate datasets. The full convolution DenseNet used deconvolutions and the connection way similar to U-net, and achieved end to end image segmentation. Furthermore, improving the Dice similarity loss function could solve the problems that the proportion of background in the prostate MRI image is much larger than that of prostate region and some pixels are difficult to be accurately identify. The experiment on PROMISE12 dataset shows that the Dice similarity coefficient is over 93.25%, and Hausdorff distance is shorter than 1.2 mm. Compared with other main methods, the proposed method is more effective and takes less time.

Key words: prostate MRI segmentation; DenseNet; full convolution network; Dice loss function

0 引言

利用医学影像确定病变区域是诊治前列腺疾病的关键步骤之一。常用的前列腺影像包括计算机断层扫描(CT)影像和磁共振成像(MRI)。CT 检查具有快速、价格低廉等优点,但是图像的对比度低;MRI 检查虽然昂贵但图像更为清晰,软组织结构对比度也较高,使得前列腺与周围组织更加容易区分,被认为是诊断前列腺癌变的最有效影像^[1]。利用 MRI 影像诊断前列腺疾病,通常由医生手动分割前列腺区域。由于这个过程不仅耗时费力,而且不同医生对目标区域的分割也存在一定的差异,所以借助计算机自动分割目标区域对诊断前列腺疾病至关重要。自动分割可以完全避免医生主观因素的影响,提高处理数据的速度,而且可重复性好。然而由于以下原因,前列腺 MRI 自动分割依然十分困难:a)不同的扫描设备产生的影像存在很大的差异;b)在 MRI 中,前列腺组织区域较小,因此有效信息也较少;c)前列腺区域与其周围组织缺乏清晰的边界。

近年来,学者们提出了一系列的前列腺图像分割方法。例如,文献[2,3]采用基于图谱的方法,达到了一定的分割效果,但是分割精度在很大程度上依赖于图谱之间的相似性。文献

[4]通过将前列腺 MRI 与包含形状等信息的模型进行匹配来确定前列腺区域的边界,该算法对伪影有较好的鲁棒性,但是需要手工标记一些边界点作为先验信息,且计算复杂。文献[5]提出了一种基于活动轮廓模型的有效方法,然而当前列腺边界模糊时,分割效果会变差,甚至失败。尽管以上这些方法在某些情形下有较好的分割效果,但它们往往需要手工标记特征或先验信息,且计算复杂度高,使得其在临床应用上受到很大的限制。最近几年,随着对卷积神经网络特征提取能力的挖掘,越来越多的深度学习技术被用于医学图像分割任务。杨金鑫等人^[6]利用卷积神经网络获得初步的分割结果,然后使用超像素聚类方法进行改进提升,但是分割效率较低。詹曙等人^[7]采用全卷积神经网络来实现端到端的前列腺分割,明显提高了分割效率,但该方法因忽略了细节特征和切片之间的空间信息,影响了分割精度。文献[8]提出了一种基于 3D 卷积的全卷积神经网络分割方法,以整个病人影像的所有切片作为输入,充分利用了切片间的空间信息,分割效果较好,但是网络计算量大,且对 GPU 有很高的要求。

本文受目前流行的深度学习模型的启发,提出了一种基于全卷积 DenseNet 的前列腺 MRI 分割方法。该方法利用迁移学

收稿日期: 2018-11-10; **修回日期:** 2019-01-17 **基金项目:** 国家自然科学基金资助项目(61571071);重庆市自然科学基金重点资助项目(cstc2017jcyjXB0037)

作者简介: 胡学刚(1965-),男,重庆人,教授,硕导,博士,主要研究方向为偏微分方程与图像处理;杨洪光(1993-),男(通信作者),河南偃师人,硕士,主要研究方向为图像处理与分析(yhgstudy@126.com)。

习的思想,将目前分类性能最好的 DenseNet 从自然图像迁移到前列腺 MRI 数据集,以提高特征提取能力和训练速度;并且采用反卷积和融合特征图的方法,以在恢复分辨率的同时融合细节信息,从而实现端到端的图像分割网络;引入并改进 Dice 相似性损失函数解决了前列腺 MRI 中背景所占比例远远大于前列腺区域,且一些像素难以准确分割等问题。

1 卷积神经网络模型

1.1 分割网络常用层

在分割网络中,常用层包括卷积层(convolution layer)^[9]、批归一化层(batch normalization, BN)^[10]、修正线性单元(rectified linear unit, ReLU)^[11]、池化层(pooling layer)和反卷积层(deconvolution layer)^[12]。卷积层具有局部感受野和参数共享的特点,通常使用卷积核与输入特征图进行卷积操作以提取特征;BN 层将卷积层的输出进行标准化,以解决梯度消失和梯度爆炸问题,并且可以提高网络的训练速度;ReLU 作为目前流行的激活函数,相较于传统的激活函数 sigmoid,不仅没有饱和区,而且可以使特征图具有稀疏性,进而减小计算量并加速网络收敛;池化层对特征图进行下采样以降低分辨率,增大感受野。这样可以增强网络对尺度变换的鲁棒性,并且有效地防止过拟合。常用的池化层有最大池化层和均值池化层。反卷积又被称为转置卷积,可以视为卷积层的逆操作,目的是作为上采样方法恢复分辨率。

1.2 DenseNet 分类网络

卷积神经网络通常用于分类任务,如 VGG-Net^[13]、GoogLeNet^[14]和 ResNet^[15]等。这些网络每一层只是从上一层或上两层的特征图中学习更高层次的特征,没有考虑之前更低层次的特征图。简言之,每一层的特征在整个网络中大多只使用一次。这种方式造成了特征图重用率低,进而导致网络的学习效率不高^[16]。虽然一些研究表明,深层次的网络拥有更好的特征提取能力^[13],但是提升网络而忽略特征图的使用率反而会造成模型退化^[15,16]。鉴于此,研究者提出了 DenseNet^[16],该网络主要由密集连接块(dense block, DB)和过渡层(transition layer, TL)组成。DB 将前面所有卷积块的特征图进行重用,即每一层的输入包含了前面所有卷积块的输出,如图 1 所示。

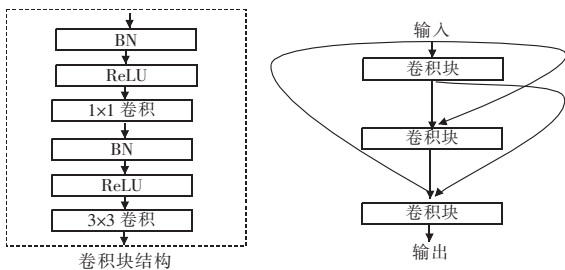


图 1 卷积块个数为 3 的密集块结构

Fig.1 Illustration of DB with 3 convolution blocks

若一个 DB 共有 L 个卷积块, x_0 作为 DB 的输入, x_l 是第 l 个卷积块的输出,可表示为

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad l=1, 2, 3, \dots, L \quad (1)$$

其中: $H_l(\cdot)$ 表示第 l 卷积块的复合函数(如图 1 所示,包含了 BN—ReLU— 1×1 卷积—BN—ReLU— 3×3 卷积), $[x_0, x_1, \dots, x_{l-1}]$ 表示将第 x_0 到 x_{l-1} 进行合并。而传统的卷积网络,第 l 个卷积块的输入仅是上一个卷积块的输出,可表示为

$$x_l = H_l(x_{l-1}) \quad (2)$$

与传统的卷积网络不同,DB 直接将各个卷积块相连接,使每个卷积块的输出都可以作为后面卷积块的输入,以保证卷积块间最大程度的信息传递。这个设计使得特征的前向传递和

梯度的后向传播更加有效,也使得特征图得到了充分利用。由于 DB 卷积块中卷积核的个数较少,在一定程度上减少了参数数量,所以这种结构使得网络训练变得更加容易^[16]。过渡层包括了 1×1 卷积层和 2×2 平均池化层,其通过 1×1 卷积减少特征图的通道数,利用池化层以降低特征图的分辨率,进一步地减少计算量并提高网络对尺度变化的鲁棒性。在文献[16]中, DenseNet 已经被证明拥有强大的特征提取能力。DenseNet 包含了 DenseNet-121、DenseNet-169、DenseNet-201 和 DenseNet-161,这些网络中卷积层和全连接层的总数分别为 121、169、201 和 161。考虑到 GPU 的情况和任务复杂度,本文采用 DenseNet-161 网络,其组成如表 1 所示。

表 1 DenseNet-161 网络结构
Tab.1 Architecture of DenseNet-161

结构	分辨率大小	DenseNet-161
输入层	224 × 224	-
卷积层 1	112 × 112	7 × 7, 96 conv, 步长 2
池化层 1	56 × 56	3 × 3 最大池化, 步长 2
密集连接块 1	56 × 56	$\begin{bmatrix} 1 \times 1, 192 \text{ conv} \\ 3 \times 3, 48 \text{ conv} \end{bmatrix} \times 6$
过渡层 1	56 × 56 28 × 28	1 × 1 conv 2 × 2 平均池化, 步长 2
密集连接块 2	28 × 28	$\begin{bmatrix} 1 \times 1, 192 \text{ conv} \\ 3 \times 3, 48 \text{ conv} \end{bmatrix} \times 12$
过渡层 2	28 × 28 14 × 14	1 × 1 conv 2 × 2 平均池化, 步长 2
密集连接块 3	14 × 14	$\begin{bmatrix} 1 \times 1, 192 \text{ conv} \\ 3 \times 3, 48 \text{ conv} \end{bmatrix} \times 36$
过渡层 3	14 × 14 7 × 7	1 × 1 conv 2 × 2 平均池化, 步长 2
密集连接块 4	7 × 7	$\begin{bmatrix} 1 \times 1, 192 \text{ conv} \\ 3 \times 3, 48 \text{ conv} \end{bmatrix} \times 24$
分类层	1 × 1	7 × 7 全局平均池化 1 000 维全连接层, softmax

在表 1 中, conv 表示 BN—ReLU—卷积的组合; $[\] \times d$ 表示 d 个卷积块。DenseNet-161 中最后两层为分类层,分类层之前的部分可称为特征提取部分,考虑到分割网络是全卷积网络,不需要分类层,因此将分类层去掉,只用前面的神经网络层(表示为 DenseNet-160)作为分割网络的特征提取部分,亦称为编码器。

2 基于全卷积 DenseNet 网络结构的新方法

如上所述,加深网络可以提高网络的学习能力,但如果忽略特征图的使用率,效果提升不显著。DenseNet 网络结构采用密集连接的方式以提高特征图的重用率,从而进一步提高网络的学习效率,在 ImageNet 分类任务上取得了杰出的成果^[16]。

然而,分割任务与分类任务不同,其要求输出的是张分割图,而不是一个一维向量。Shelhamer 等人^[17]提出了将 VGG 分类网络的全连接层卷积化,形成全卷积神经网络,使输出预测图与输入图像的尺寸一致,实现了端到端的自然图像语义分割。此后,全卷积神经网络被认为是图像分割的基础网络。文献[18]提出了 U-Net 图像分割网络,该网络同样是全卷积神经网络,其采用编/解码结构,编码器用于提取特征,解码器用于恢复分辨率,并且采用拼接的方式将编码器和解码器进行连接,以融合前面特征层所学习到的细节信息。该网络在许多医学图像分割任务上取得了不错的结果。

本文受 DenseNet 和 U-Net 等神经网络的启发,提出了一种基于全卷积 DenseNet 网络结构的前列腺 MRI 分割方法。该方法包含训练和测试阶段,具体算法如图 2 所示。在训练阶段,采用交叉验证法,从 50 个病人预处理后的前列腺 MRI 中,随机抽取 5 例作为验证集,剩余的为训练集,并对训练集和验证

集作归一化预处理;加载 DenseNet-160 预训练模型的权重到全卷积 DenseNet 网络的编码器部分;将训练集的前列腺 MRI 及其标注图送入全卷积 DenseNet 网络,经过前向传播,并利用改进的 Dice 损失函数和梯度下降法不断地更新模型的参数;用验证集为模型训练提供监督指导,以保存最佳模型。在测试阶段,对测试图像作预处理后输入到保存的最佳模型中,经过前向传播,最终得到测试图像的分割结果图。

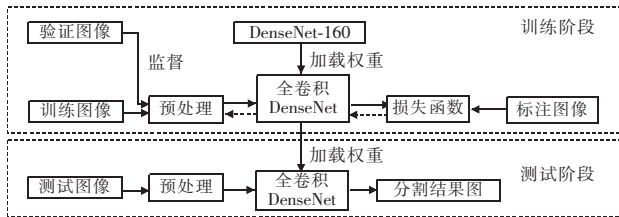


图 2 本文方法流程
Fig. 2 Diagram of the proposed method

2.1 全卷积 DenseNet

本文提出的全卷积 DenseNet 网络结构如图 3 所示。对于切片 i , 将其相邻的两张切片 $(i-1, i+1)$ 与之一起输入网络, 网络的输出为中间切片 i 的分割结果。换言之, 相邻的两张切片仅作为辅助信息。这样可以使网络学习到切片间的空间信息, 能更准确地分割出中间切片的前列腺部分, 而且相较于文献[8]中以整个病人影像的所有切片作为输入和采用 3D 卷积的方法, 计算量更小并且节省了 GPU 资源。

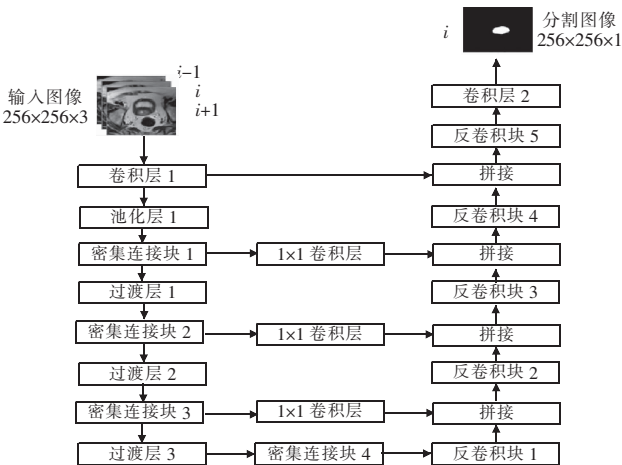


图 3 全卷积 DenseNet 网络结构
Fig. 3 Illustration of fully convolutional DenseNet

类似于 U-Net 结构,全卷积 DenseNet 网络由编码器和解码器组成。卷积层 1 至密集连接块 4 组成的 DenseNet-160 为本文分割网络的编码器部分,具体组成结构如表 1 所示。较 U-Net 中编码器部分仅有 10 个卷积层不同,全卷积 DenseNet 网络的编码器部分有 160 个卷积层。由于网络深度越深提取特征能力越强^[13],所以本文分割网络具有很强的前列腺特征提取能力。此外,如 1.2 节中所述,该编码器的密集连接方式不仅可以提高特征图的重用率,而且可以使梯度传递更加有效,以防止梯度消失。解码器部分包含了反卷积块 1~5 和卷积层 2。反卷积块由 1×1 卷积层、 3×3 卷积层和反卷积层组成,且后面都紧接 BN 和 ReLU 激活函数。同一个反卷积块中所有卷积层的卷积核个数一致。反卷积块 1~5 中的卷积核个数分别为 768、384、96、96 和 64。在反卷积块中, 1×1 卷积层的卷积核个数远远小于其输入特征图的通道数,以降低输出特征图的通道数,进而使其后面 3×3 卷积核的参数大大减少。因为卷积神经网络的浅层特征图得到的是局部细节信息,深层特征图学习到的是全局语义信息^[17],在解码器进行恢复分辨率的

时候,将两者进行融合以学习上下文信息可以提高分割精度^[18]。为此,本文采用类似 U-Net 的拼接方式将编码器与解码器进行连接,如图 3 所示。由于密集连接块的输出具有很高的通道数,所以本文在密集连接块与解码器中间插入了 1×1 卷积以降低通道数,这种连接方式与 U-Net 直接将编码器和解码器连接不同。 1×1 卷积是轻量级的,且没有改变输出的感受野,常常被用来改变特征图的通道数^[14-16]。反卷积块 2~5 中的 3×3 卷积层可以对拼接后的特征图进行融合和学习上下文信息。反卷积层作为上采样方法的一种,不仅可以增大特征图的分辨率,并且是可学习的,在很多分割网络中被应用^[17]。卷积层 2 为 1 个 1×1 卷积,并采用 sigmoid 激活函数使输出在 $(0, 1)$, 表示属于正样本(前列腺区域)的概率。该解码器不仅可以在恢复分辨率的同时有效地学习上下文信息,而且 1×1 卷积也大大降低了参数量。

2.2 改进的 Dice 损失函数

全卷积 DenseNet 网络的输出为一个与原始图像分辨率一致的分类概率图,每一个像素点的值表示属于正样本的可能性。根据损失函数计算分类概率图与真实标签间的误差,优化损失函数,通过反向传播算法迭代权重,最终获得收敛的网络模型。与大多数自然图像不同,在前列腺 MRI 中,属于前列腺的区域非常小,而背景区域远远大于目标区域。因此,在网络训练的过程中,模型更容易偏向学习背景的特征,导致在检测时很难检测到目标区域。为了使网络更加着重学习目标特征,某些学者根据 Dice 相似性系数使用了损失函数 L_{Dice} ^[7],表示为

$$L_{Dice} = 1 - \frac{\sum_i p_i g_i}{\sum_i p_i^2 + \sum_i g_i^2} \quad (3)$$

其中: $p_i \in (0, 1)$ 和 $g_i \in \{0, 1\}$ 分别表示分类概率图和真实标签中第 i 个像素点的值。等式右边的第二项可用来表示分割结果与真实标签间的相似性。在训练过程中,使用小化 L_{Dice} 损失函数以达到最大化相似性的目的。然而该损失函数并不能使网络着重学习那些比较难分割的像素特征。注意到,当 $g_i = 1$ 时, $p_i g_i$ 才为非零值。换言之,损失函数的分子只与正样本像素点有关。在正样本像素点位置, p_i 越小,表示网络识别它为正样本的概率很低,说明该像素点是困难样本,应该加大其在损失函数中的权重,使网络着重学习困难样本,以进一步提高分割准确度。为此,本文提出一种自适应权重的 Dice 损失函数 $L_{adaDice}$,即

$$L_{adaDice} = 1 - \frac{\sum_i p_i^2 g_i}{\sum_i p_i^2 + \sum_i g_i^2} \quad (4)$$

容易发现,将 p_i, g_i 乘以 p_i ,当 p_i 越小时,等式右边的第二项就越小, $L_{adaDice}$ 就越大,由此提高困难样本在损失函数中的权重。实验表明,采用自适应权重的 Dice 损失函数比一般损失函数有更好的性能。

3 实验与结果分析

3.1 数据集

在深度学习中,往往需要大量的数据来训练模型。尽管本文利用迁移学习的思想,将在 ImageNet 数据集上预训练好的 DenseNet-160 加载到全卷积 DenseNet 可以提高训练速度,但是仍然需要大量的前列腺 MRI 图像来提高模型的鲁棒性。本文采用网上公开的 PROMISE12 数据集,该数据集包含了训练集和测试集,分别有 50 和 30 个病人样本。每个样本的前列腺 MRI 图像包含了 20~40 个切片,训练集还有对应的分割标签图。由于 PROMISE12 数据集的图像数量不大,所以本文采取随机剪裁、尺度变换、翻转等方式扩充数据集。

为了更好地训练模型,本文采用以下预处理步骤:a)根据分割标签图,从训练集中去除不包含前列腺区域的切片,以减少负样本的数目;b)对训练图像进行归一化操作,操作定义为

$$I = \frac{X - \mu}{\sigma} \quad (5)$$

其中: X 表示原始图像像素; μ 表示原始图像像素的均值; σ 表示原始图像像素的标准差; I 表示归一化后的图像。

3.2 实验设置

本文采用 TensorFlow^[19] 开源深度学习库,使用 NVIDIA 1080Ti 显卡进行加速。训练时,加载预训练好的 DenseNet-160,用于初始化全卷积 DenseNet 中编码器的网络参数,其他层采用随机初始化的方法。本文使用 Adam 优化算法训练模型,初始学习率为 0.001,使用 poly 学习率衰减策略^[20],并且采用 l_2 正则化(衰减因子为 0.0005)以降低过拟合情况。

测试时,输入一张任意大小且归一化后的前列腺 MRI 图像到训练好的网络中,通过前向传播输出一张分割概率图,并根据 0.5 阈值进行二值化,最后生成分割预测图。

3.3 实验结果

本文实验采用 Dice 相似性系数(DSC)和 Hausdorff 距离(HD)来定量评价不同方法的分割效果。用 P 、 G 分别表示预测前列腺区域和真实前列腺区域所包含的点集,则 DSC 定义为

$$DSC(P, G) = \frac{2|P \cap G|}{|P| + |G|} \quad (6)$$

DSC 越高表示两个区域越相似。HD 可反映出两个区域点集的最大差异,定义为

$$HD(P, G) = \max \left(\begin{array}{l} \max_{p_i \in P} \min_{g_j \in G} \|p_i - g_j\| \\ \max_{g_j \in G} \min_{p_i \in P} \|p_i - g_j\| \end{array} \right) \quad (7)$$

HD 越小表示分割准确率越高。

表 2 展示了不同方法的定量评价结果。其中文献[21~23]为传统方法,文献[7,24]和本文算法均为深度学习方法。可以看出,基于深度学习方法的分割性能要优于传统方法。堆叠 ISA 方法^[24]尽管采用了深度学习框架,但不是端到端的分割网络,时间复杂度大。反卷积网络方法^[7]采用全卷积神经网络,可以直接输出分割预测图,但是其网络结构简单,特征提取能力不足,且在恢复分辨率时没有融合前面卷积层所学习到的细节信息。本文方法采用所提出的全卷积 DenseNet 结构,提取特征能力强,并且可以充分学习上下文信息,自适应的 Dice 损失函数可以促使网络着重学习前景区域和困难样本,进而提高分割效果。虽然本文网络结构比反卷积网络方法复杂,但是不需要对图像作对比度增强预处理,因此运行时间也更小,达到实际临床需求。

表 2 不同方法分割结果的定量比较

Tab. 2 Quantitative comparison of segmentation results

方法	DSC/%	HD/mm	时间/min
Tian ^[21]	83.4	9.3	4
活动轮廓方法 ^[22]	86	9.5	6~8
Mahapatra ^[23]	92.1	5.5	23
堆叠 ISA 方法 ^[24]	86.7	1.9	1.4
反卷积网络方法 ^[7]	89.75	1.3	0.3
本文方法	93.25	1.1	0.2

图 4 给出了本文方法的实验分割结果与真实标注图的对比。蓝色和红色分别表示本文方法分割结果和金标准的轮廓(见电子版)。可以看出,图中的六张图具有对比度低和前列腺区域边界模糊的特点,本文方法仍能较好地分割出前列腺区域。

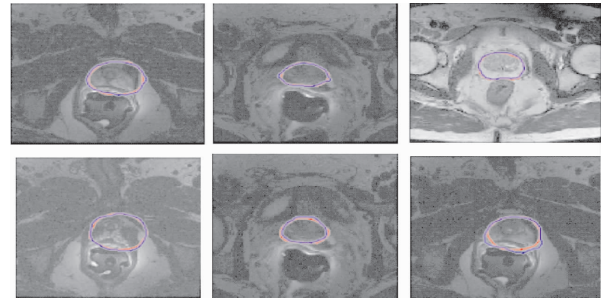


图 4 前列腺 MRI 分割结果

Fig. 4 Segmentation results of prostate MRI

4 结束语

本文受目前流行的深度学习模型的启发,提出了一种基于全卷积 DenseNet 的前列腺 MRI 分割方法。该方法利用迁移学习的思想,将目前分类性能较好的 DenseNet 从自然图像迁移到前列腺数据集,以提高特征提取能力和训练速度。并且采用反卷积和融合特征图的方法,以在恢复分辨率的同时学习上下文信息,从而实现端到端的图像分割网络。引入并改进 Dice 相似性损失函数以解决前列腺 MRI 影像中背景所占比例远远大于前列腺区域且一些正样本难以准确分割等问题。通过实验分析,本文所提出的方法相较于目前的主要方法,分割性能更好,所耗时间更少。

参考文献:

- [1] Sharp G, Fritscher K D, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy [J]. *Medical Physics*, 2014, 41(5): 050902.
- [2] Klein S, Der Heide U A V, Raaymakers B W, et al. Segmentation of the prostate in MR images by atlas matching [C]//Proc of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Piscataway, NJ: IEEE Press, 2007: 1300-1303.
- [3] Martin S, Troccaz J, Daanenc V. Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model [J]. *Medical Physics*, 2010, 37(4): 1579-1590.
- [4] Heimann T, Meinzer H P. Statistical shape models for 3D medical image segmentation: a review [J]. *Medical Image Analysis*, 2009, 13(4): 543-563.
- [5] Toth R, Madabhushi A. Multifeature landmark-free active appearance models: application to prostate MRI segmentation [J]. *IEEE Trans on Medical Imaging*, 2012, 31(8): 1638-1650.
- [6] 杨鑫, 杨辉华, 李灵巧, 等. 结合卷积神经网络和超像素聚类的细胞图像分割方法 [J]. *计算机应用研究*, 2018, 35(5): 1569-1572, 1577. (Yang Jinxin, Yang Huihua, Li Lingqiao, et al. Cell image segmentation method based on convolution neural network and super pixel clustering [J]. *Application Research of Computers*, 2018, 35(5): 1569-1572, 1577.)
- [7] 詹曙, 梁植程, 谢栋栋. 前列腺磁共振图像分割的反卷积神经网络方法 [J]. *中国图象图形学报*, 2017, 22(4): 516-522. (Zhan Shu, Liang Zhicheng, Xie Dongdong. Deconvolutional neural network for prostate MRI segmentation [J]. *Journal of Image and Graphics*, 2017, 22(4): 516-522.)
- [8] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C]//Proc of the 4th International Conference on 3D Vision. Piscataway, NJ: IEEE Press, 2016: 565-571.
- [9] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [10] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. (2015-03-02). <https://arxiv.org/abs/1502.03167>. (下转第 1894 页)

地得到大量的训练样本,而无须人工标注。同时,由于转换跳过了中间的文字表示环节,可以避免文字的发音歧义。实验表明,本方法基于 LSTM 网络和 LPC10 声码器技术,可以由嘴唇视频直接重建较为清晰易懂的语音。然而,本文测试的数字和礼貌用语尚属简单有限的词汇和短语,而且训练样本和测试样本采集于同一个人的视频,本方法尚有较大的继续研究空间。在后续的工作中,将尝试较大的词汇集和提高在不同人之间的泛化能力。

参考文献:

- [1] Assael Y M, Shillingford B, Whiteson S, *et al.* LipNet: end-to-end sentence-level lipreading [EB/OL]. (2016-12-16). <https://arxiv.org/abs/1611.01599>.
- [2] Fernandez-Lopez A, Sukno F M. Survey on automatic lip-reading in the era of deep learning [J]. *Image and Vision Computing*, 2018, 78(10): 53-72.
- [3] Thein T, San K M. Lip localization technique towards an automatic lip reading approach for Myanmar consonants recognition [C]//Proc of International Conference on Information and Computer Technologies. Piscataway, NJ: IEEE Press, 2018: 123-127.
- [4] Goldschien A J, Garcia O N, Petajan E. Continuous optical automatic speech recognition by lipreading [C]//Proc of the 28th Asilomar Conference on Signals, Systems and Computers. Piscataway, NJ: IEEE Press, 1994: 572-577.
- [5] Wand M, Koutník J, Schmidhuber J. Lipreading with long short-term memory [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2016: 6115-6119.
- [6] Almajai I, Cox S, Harvey R, *et al.* Improved speaker independent lip reading using speaker adaptive training and deep neural networks [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2016: 2722-2726.
- [7] Cooke M, Barker J, Cunningham S, *et al.* An audio-visual corpus for speech perception and automatic speech recognition [J]. *Journal of the Acoustical Society of America*, 2006, 120(5): 2421-2424.
- [8] Mroueh Y, Marcheret E, Goel V. Deep multimodal learning for audio-visual speech recognition [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2015: 2130-2134.
- [9] Petridis S, Pantic M. Deep complementary bottleneck features for visual speech recognition [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2016: 2304-2308.
- [10] Oord A V D, Dieleman S, Zen Heiga, *et al.* WaveNet: a generative model for raw audio [EB/OL]. (2016-09-19). <https://arxiv.org/abs/1609.03499>.
- [11] Oord A V D, Li Yazhe, Babuschkin I, *et al.* Parallel WaveNet: fast high-fidelity speech synthesis [EB/OL]. (2017-11-28). <https://arxiv.org/abs/1711.10433>.
- [12] 黄源浩, 刘龙, 肖振中, 等. 自适应唇舌交互方法以及交互装置: 中国, CN106504751A [P]. 2016-11-29. (Huang Yuanhao, Liu Long, Xiao Zhenzhong, *et al.* Adaptive lip-language interaction methods and devices: China, CN106504751A [P]. 2016-11-29.)
- [13] 焦友明, 刘念邱. 一种基于唇舌采集和语音拾取的多语言翻译器: 中国, CN107610703A [P]. 2018-01-19. (Jiao Youming, Liu Nianqiu. A multilingual translator based on lip language acquisition and voice pickup: China, CN107610703A [P]. 2018-01-19.)
- [14] King D E. Dlib-ml: a machine learning toolkit [J]. *Journal of Machine Learning Research*, 2009, 10(3): 1755-1758.
- [15] Sagonas C, Tzimiropoulos G, Zafeiriou S, *et al.* 300 faces in-the-wild challenge: the first facial landmark localization challenge [C]//Proc of IEEE International Conference on Computer Vision Workshops. Piscataway, NJ: IEEE Press, 2013: 397-403.
- [16] 张雪英, 贾海蓉. 语音与音频编码 [M]. 西安: 西安电子科技大学出版社, 2011: 82-87. (Zhang Xueying, Jia Hairong. Speech and audio coding [M]. Xi'an: Xi'an University of Electronic Science and Technology Press, 2011: 82-87.)
- [17] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [18] Cho K, Van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. (2014-09-03). <https://arxiv.org/abs/1406.1078>.
- [19] Easton R D, Basala M. Perceptual dominance during lipreading [J]. *Perception & Psychophysics*, 1982, 32(6): 562-570.
- [20] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30). <https://arxiv.org/abs/1412.6980>.
- [21] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines [C]//Proc of the 27th International Conference on Machine Learning. 2010: 807-814.
- [22] Zeiler M D, Taylor G W, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning [C]//Proc of International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2011: 2018-2025.
- [23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10). <https://arxiv.org/abs/1409.1556>.
- [24] Szegedy C, Wei Liu, Yangqing Jia, *et al.* Going deeper with convolutions [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015: 1-9.
- [25] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 770-778.
- [26] Huang Gao, Liu Zhuang, Van Der Maaten L, *et al.* Densely connected convolutional networks [EB/OL]. (2018-01-28). <https://arxiv.org/abs/1608.06993>.
- [27] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [EB/OL]. (2016-05-20). <https://arxiv.org/abs/1605.06211>.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.
- [29] Abadi M, Barham P, Chen Jianmin, *et al.* TensorFlow: a system for large-scale machine learning [EB/OL]. (2016-05-31). <https://arxiv.org/abs/1605.08695>.
- [30] Liu Wei, Rabinovich A, Berg A C. ParseNet: looking wider to see better [EB/OL]. (2015-11-19). <https://arxiv.org/abs/1506.04579>.
- [31] Tian Zhiqiang, Liu Lizhi, Fei Baowei. A fully automatic multi-atlas based segmentation method for prostate MR images [C]//Proc of Medical Imaging: Image Processing. 2015: 1067-1073.
- [32] Kirschner M, Jung F, Wesarg S. Automatic prostate segmentation in MR images with a probabilistic active shape model [EB/OL]. (2013-09). <https://www.researchgate.net/publication/259965252>.
- [33] Mahapatra D, Buhmann J M. Prostate MRI segmentation using learned semantic knowledge and graph cuts [J]. *IEEE Trans on Biomedical Engineering*, 2014, 61(3): 756-764.
- [34] Liao Shu, Gao Yaorong, Oto A, *et al.* Representation learning: a unified deep learning framework for automatic prostate MR segmentation [C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2013: 254-261.