

基于Unet的阿尔兹海默症病灶分割模型

徐 超, 王卓薇, 刘晓东

广东工业大学, 计算机学院, 广东 广州

收稿日期: 2021年12月25日; 录用日期: 2022年1月20日; 发布日期: 2022年1月27日

摘 要

由于人脑的结构比较复杂, 所以如何准确的提取分割部位的特征是医学图像语义分割的一个关键点。本文使用Unet网络作为基本框架, 并在特征提取过程中引入了注意力机制CBAM, 在Unet的Skip-connection过程中引入了SENet。通过这些方法从全局增强分割部位的语义信息并降低其他部位的语义信息。本文使用相同的设备和数据分别在Unet、Unet + CBAM、Unet + SENet以及本文提出的模型进行了实验, 得到对应的平均Dice分别对应为0.680、0.703、0.706、0.736。实验表明, 本文提出的方法对阿尔兹海默症病灶分割的性能有显著提升。

关键词

医学图像, 语义分割, CBAM, SENet, Unet

Unet-Based Model of Alzheimer's Disease Lesion Segmentation

Chao Xu, Zhuowei Wang, Xiaodong Liu

School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong

Received: Dec. 25th, 2021; accepted: Jan. 20th, 2022; published: Jan. 27th, 2022

Abstract

Because the structure of the human brain is relatively complex, how to accurately extract the features of the segmented part is a key point in the semantic segmentation of medical images. This paper uses Unet network as the basic model framework, and introduces the attention mechanism CBAM in the feature extraction process, and utilizes SENet in the Skip-connection process of Unet. Through these methods, the semantic information of segmented parts is globally enhanced and the semantic information of other parts is reduced. We compare the proposed method with Unet, Unet + CBAM, and Unet + SENet on the same equipment and data, and the corresponding average

Dice is 0.736, 0.680, 0.703, 0.706, respectively. Experiments show that the proposed method can significantly improve the performance of Alzheimer's disease lesion segmentation.

Keywords

Medical Image, Semantic Segmentation, CBAM, SENet, Unet

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

阿尔兹海默症(Alzheimer's Disease, AD)是一种以记忆功能和认知功能进行性退化为特征的临床综合征,是导致痴呆症的主要原因,也是导致人类死亡的一大原因。AD 的潜伏期较长,其临床表现也会随着时间的推移而慢慢恶化,它通常隐匿起病,持续进行性发展,临床上表现为记忆障碍、失语、失用、失认以及视空间能力损害等一系列认知功能减退和非认知性神经精神症状[1],而要在相关症状出现之前准确诊断患病风险并不容易。磁共振成像(MRI)是医学专业人员用来诊断神经和脑相关疾病的重要工具。由于它具有非侵入性、可以提供高空间分辨率的图像等特点,成为了脑结构分析的首选方法。研究发现,阿尔兹海默患者的海马体、颞叶以及额叶等部位会比正常衰老的老人萎缩的程度更大。所以,如果我们可以直接的观测到这些组织,那么对于医生的临床诊断将会有很大的帮助。

常用的医学图像分割方法可以分为传统型的和深度学习型的,传统型的有随机森林(Random Forest, RF)、遗传算法(Genetic Algorithm, GA)、支持向量机(Support Vector Machines, SVM)等方法。深度学习的主要有编码-解码器、模型压缩、注意力机制、图像合成等方法。而随着近些年对深度学习研究的加深,越来越多的学者使用深度学习的方法对医学图像的语义分割展开研究。由于每种方法对解决的问题具有一定的针对性,所以对每种方法的评价都需要结合实际需要解决的问题来进行。而我们的工作是对阿尔兹海默症患者的脑 MRI 图像,针对已有的数据集特点以及脑部结构的复杂性,上述的这些方法要么在处理复杂的不连续结构中难以获得满意的效果,要么就是只能进行单个病灶的识别,很难达到一个令人满意的效果。因此,在现有的 MRI 图像分析方法中应该进一步关注如何对医学图像这样一个特殊的数据进行特征提取,如何对脑部这样一个复杂的环境进行分析,如何使用少量的带标注的数据集也能有一个很好的效果,这些都是我们所要考虑的问题。

针对脑部医学图像的特征,本文提出了一种以 Unet 为基础适用于阿尔兹海默病灶分割的模型。本文的主要贡献有以下两点:

- 1) 本文将注意力机制运用在特征提取的过程中,改善了因脑部图像结构复杂,目标体积小,而导致特征提取性能低的问题。
- 2) 本文在跳跃连接过程中引入了 SENet 进行处理后再进行连接,用以增强低层语义和高层语义的连接效果。并使模型可以更好的进行多尺度预测和为分割提供更加精细的特征。

2. 相关工作

卷积神经网络(Convolutional Neural Networks, CNN)在图像分割、预测和分类等应用中得到了广泛的应用。Long 等人于 2015 年时提出了一种用于像素级标记的完全卷积神经网络模型(FCN)新的分割模型

[2], 他们采用反卷积对最后一个卷积层的特征图进行上采样, 使它恢复到输入图像相同的尺寸, 从而可以对每个像素都产生了一个预测。同时, 他们使用 Skip-connection 将底层语义和高层语义进行融合, 使得到的结果更加平滑。与之类似的还有[3]中的 U-Net 网络模型也使用了 Skip-connection, 该模型是基于编码-解码且完全对称的网络模型。不同的是 FCN 中 Skip-connection 联合是通过对应像素的求和, 而 U-Net 中的 Skip-connection 联合是通过对其通道的拼接。在当时, U-Net 通过对细胞和细胞膜图像的分割取得了极好的效果而得到了大家的关注。以至于, 后面许多对医学图像的分割网络中, 很大一部分都是基于 U-Net 进行改进而得到的。3DUnet [4]的诞生给医学影像分割, 特别是那些需要在三维影像上进行语义分割的任务带来了极大的帮助, 很大程度上解决了 3D 图像分割时需要将一个个切片输入模型进行训练的麻烦, 也大幅度提升了训练效率。Zhou 等人提出了一种在解码器网上聚合不同语义尺度的特征, 从而产生高度灵活的特征融合的网络模型 Unet++ [5]。Huang 等人提出了一种全尺寸连接的 U-Net 模型 Unet3+ [6]。Fabian 等人使用 nnUnet 在脑肿瘤分割中获得较好的结果[7]。H-DenseUnet 可用于分割肝脏和肝肿瘤[8]。Alom 等人提出了一个递归 U-Net 模型和一个递归剩余 U-Net 模型, 分别称为 RU-Net 和 R2U-Net。所提模型结合了 U-Net、剩余网络和循环卷积神经网络。其中, 利用循环卷积层进行特征累积, 可以更好地表示分割任务的特征, 具有更好的图像分割性能[9]。

注意力机制在近些年中也渐渐的被引入到计算机视觉领域中了, Wang 等人在深度残差网络(ResNet)中增加了一个用于图像分类的注意模块[10], 他们提出的注意力模块由几个编解码层组成。Nie 等人设计了一种注意力模型, 可以更准确的从 MRI 图像中分割前列腺[11]。Sinha 等人提出了一种基于多层注意力腹部器官分割的方法[12]。此外, 对抗性训练在近些年得到了迅速的发展, 不少学者也将其应用到了医学图像语义分割中。Khosravan 等人提出了一种用于分割 CT 图像中胰腺的对抗性训练框架[13]。Son 等人使用 GANs 算法来分割视网膜图像[14]。Xue 等人基于对抗性网络从 MRI 图像中分割出了脑肿瘤[15]。

3. 本文方法

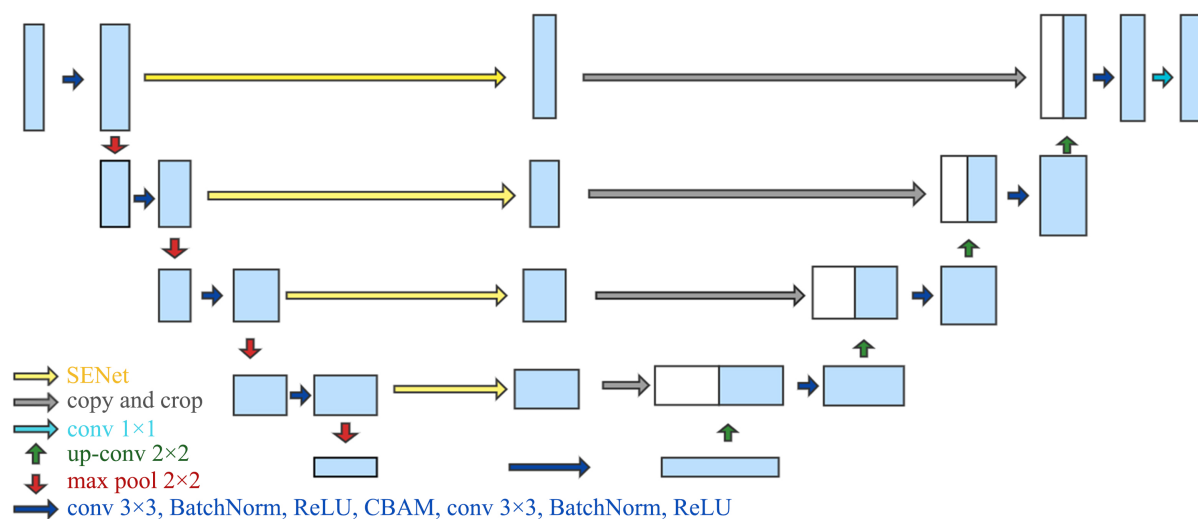


Figure 1. The architecture of the proposed method

图 1. 网络模型图

由于我们数据集的数量较少, 以及对模型参数数量的控制, 本文选择了在 U-Net 的网络模型上进行改进。考虑到人体脑部中的组织较多, 结构也相对复杂, 且分割的部位中有些部位体积较小。因此我们在模型中加入了两种注意力机制: CBAM [16]和 SENet [17]。其中, CBAM 用于特征提取的时候, 而 SENet 用

于网络 Skip-connection 的时候。使得我们的网络模型在特征提取的时候，更加多的去关注我们需要分割的目标部位，而不是背景或其他不需要关注的部位。同时在 Skip-connection 的时候，可以尽可能的通过通道的特征相应去全局的关注我们的目标特征。我们的网络模型概述图如图 1 所示。

3.1. 特征提取过程中使用 CBAM

CBAM 由 Woo 等人于 2018 年提出，CBAM 通过空间注意力和通道注意力分别进行空间和通道上的注意力机制。同时它也是一种轻量的注意力机制，可以插入到整个网络的卷积模块中，实现端到端的同步训练。针对我们的数据特点，我们在每次特征提取的时候加入了 CBAM，改进后特征提取模块如图 2 所示。

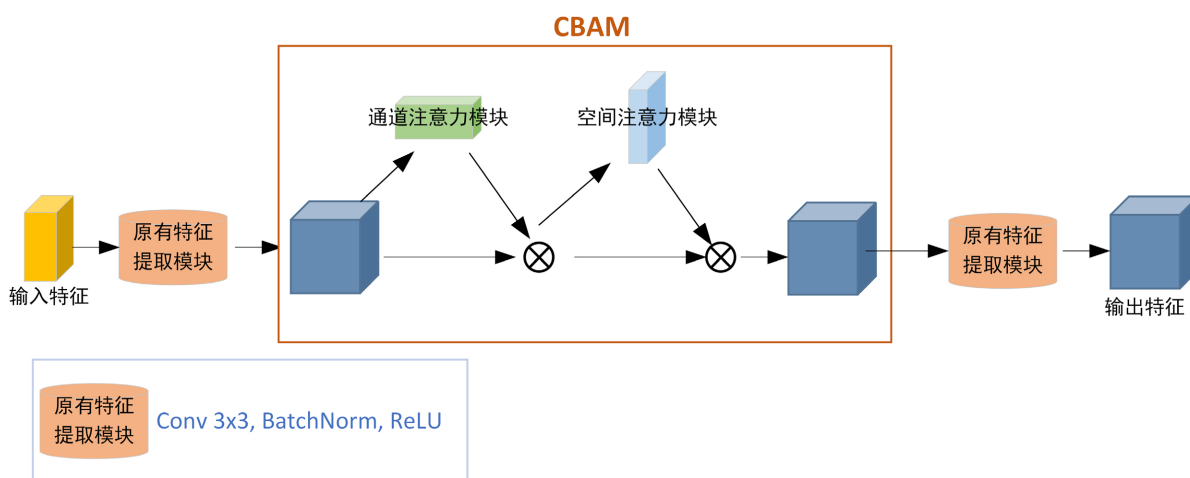


Figure 2. Feature extraction module incorporating CBAM

图 2. 包含 CBAM 后的特征提取模块

输入特征在经过原有的两个特征提取模块中间还要经过 CBAM 模块，我们可以看到 CBAM 中主要包含两个模块：通道注意力模块和空间注意力模块。这两个模块分别在特征的通道和空间两个方面使分割部位获得更多的关注。当给定一个特征图 $F \in \mathbb{R}^{C \times H \times W}$ 作为输入，该特征图会依次经过 CBAM 中的通道注意力模块和空间注意力模块，并且在经过两个模块后都会与进入模块前的特征进行逐个元素相乘，同时也会去广播相应的关注特征：会沿着空间维度广播通道维度的关注特征，也会沿通道维度广播空间维度的特征，整个过程如公式(1)所示：

$$\begin{aligned} F1 &= M_c(F) \otimes F, \\ F2 &= M_s(F1) \otimes F1 \end{aligned} \quad (1)$$

在公式(1)中， F 表示输入特征， $M_c(\cdot)$ 表示通道注意力模块(其中的 \cdot 表示输入的值)。 \otimes 表示按元素逐个相乘， $F1$ 表示经过通道注意力模块后与 F 进行 \otimes 运算后得到的特征。同理， $M_s(\cdot)$ 表示空间注意力模块， $F2$ 表示经过空间注意力模块处理后与 $F1$ 进行 \otimes 运算后得到的特征，同时也是整个 CBAM 模块处理后得到的特征。

在通道注意力模块中首先使用了平均池化和最大池化操作来聚合特征要素中的空间信息，生成了两种不同的上下文描述特征 $F_{avg}^c \in \mathbb{R}^{C \times 1 \times 1}$ (经过平均池化层得到)和 $F_{max}^c \in \mathbb{R}^{C \times 1 \times 1}$ (经过最大池化层得到)。然后，这两个描述特征被输入到多层感知机(MLP)中处理后按元素进行逐个相加后，最后经过激活函数得

到通道注意力映射图 $M_c \in \mathbb{R}^{C \times l}$ 。整个过程如公式(2)所示:

$$\begin{aligned} M_c(F) &= \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{max}^c\right)\right)\right) \end{aligned} \quad (2)$$

其中 σ 表示 Sigmoid 函数, W_0 和 W_1 是 MLP 的权重, 两个输入 F_{avg}^c 和 F_{max}^c 是共享使用这两个权重的。

与通道注意力模块相似, 空间注意力模块中对输入的特征也要先进行最大池化和平均池化。不同的是, 空间注意力模块中是对通道维度来进行聚集的, 池化后也得到两种描述特征 $F_{max}^s \in \mathbb{R}^{l \times H \times W}$ (经过最大池化层得到) 和 $F_{avg}^s \in \mathbb{R}^{l \times H \times W}$ (经过平均池化层得到), 并将这两个描述特征按照通道维度拼接起来。然后使用卷积核进行卷积操作, 并将卷积后的特征使用激活函数 Sigmoid 函数激活便得到空间注意力映射图 $M_s \in \mathbb{R}^{l \times H \times W}$, 具体的过程如公式(3)所示:

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7 \times 7}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) \\ &= \sigma\left(f^{7 \times 7}\left(\left[F_{avg}^s; F_{max}^s\right]\right)\right) \end{aligned} \quad (3)$$

其中 σ 表示 Sigmoid 函数, $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积操作, $\left[F_{avg}^s; F_{max}^s\right]$ 表示将 F_{avg}^s 和 F_{max}^s 两个特征按照通道维度进行拼接。

3.2. Skip-Connection 过程中使用 SENet

SENet (Squeeze-and-Excitation Networks) 的目的是希望显式地建模特征通道之间的相互依赖关系, 采用了特征重标定策略。简单的说就是通过学习的方式来自动获取到每个特征通道的重要程度, 然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征。SENet 首先会对卷积得到的特征图进行 Squeeze 操作, 得到通道级的全局特征, 然后对全局特征进行 Excitation 操作, 学习各个通道间的关系, 也得到不同通道的权重, 最后乘以原来的特征图得到最终特征。本质上, SENet 是在通道维度上做注意力机制或者门控制操作, 值得一提的是 SENet 是通用的, 可以嵌入到现有的网络框架中。

我们可以看出在 SENet 中, 比较重要的两个操作是 Squeeze 和 Excitation 操作。由于传统的卷积只是在一个局部空间内进行操作, 很难得到足够的信息来提取各个通道之间的关系。因此, 在 SENet 中提出了 Squeeze 操作, 将一个通道上整个空间的特征编码为一个全局的特征, 采用全局平均池化来实现, 每个通道上的具体操作如公式(4)所示:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (4)$$

其中 H 和 W 分别表示输入特征在空间维度的高和宽, $u_c(i, j)$ 表示对应点的像素值, z_c 为通道 u_c 经过 Squeeze 操作后的对应输出。Squeeze 操作得到了全局描述特征, 接下来需要另外一种运算来捕获通道之间的依赖关系。这个操作需要满足两个条件: 第一它要灵活, 它要可以学习到各个通道之间的非线性关系; 第二点是学习的关系而不是互斥的, 因为这里允许有多个通道的特征, 而不是 one-hot 形式。因此, 这里采用 sigmoid 形式的门控制机制, 具体过程如公式(5)所示:

$$s = F_{ex}(z, W) = \sigma\left(g(z, W)\right) = \sigma\left(W_2 \delta\left(W_1 z\right)\right) \quad (5)$$

其中 $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$, $W_2 \in \mathbb{R}^{\frac{C \times C}{r}}$ (r 是降维超参数, 为了减少参数数量和计算量); δ 表示 ReLU 函数, 可以使结果获得更多的非线性特征; σ 表示 Sigmoid 函数。经过 Squeeze 和 Excitation 操作后, 最后还需要进行一

个 F_{scale} 操作，该操作具体如公式 6 所示：

$$X'_c = F_{scale}(u_c, s_c) = s_c u_c \quad (6)$$

其中特征映射 $u_c \in \mathbb{R}^{H \times W}$ ， s_c 为其对应的标量，两者相乘和得到了对应结果 X'_c 。

其实整个操作过程可以看作学习到了各个通道的权重系数，从而使得 Skip-connection 后进行上采样时从全局的角度对各个通道的特征更有辨别能力。

4. 实验与分析

4.1. 数据集

本次实验中的数据集为同时含有额叶、颞叶和海马体三个部位的人脑切片。由于数据标注困难以及同时含有三个部位的切片较少，所以我们一共只获得了 136 张数据图片，并从中选取了 100 张图片作为训练集，18 张图片作为验证集剩下的 18 张作为测试集。

4.2. 评价指标

在医学图像分割中，常用的评价指标主要有 Dice 和 Iou 两种，本次实验主要是采用 Dice 系数来作为评价指标。Dice 系数的定义如公式(7)所示：

$$\text{Dice}(X, Y) = \frac{2TP}{2TP + FP + FN} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7)$$

Dice 系数本质上是模型的预测输出和标签之间重叠的度量系数，其范围为 0~1，为 1 时表示完全重叠。

4.3. 实验结果

为了验证我们的模型的有效性，我们使用相同的数据集在同一台设备中进行了消融实验。实验结果如表 1 所示。

Table 1. Results of ablation study

表 1. 消融实验结果

模型	结果(Dice)			
	额叶	颞叶	海马体	平均
Unet	0.814	0.813	0.413	0.680
Unet + CBAM	0.829	0.814	0.466	0.703
Unet + SENet	0.851	0.817	0.450	0.706
our	0.870	0.827	0.511	0.736

从实验结果中我们可以看出，我们在原本的 Unet 网络模型中加入了 CBAM 和 SENet 是有效果的，并且两个的效果有一定的叠加作用。从部位来看，不管是使用哪个模型，额叶的效果都是最好的，颞叶的效果相比额叶的效果要稍微低一点，但是海马体的分割效果和前两者相比就相差了很多，这主要是因为海马体的体积较小，特征提取效果偏差。为了更好的观察各个模型之间效果的对比，我们将各个模型的输出进行了可视化操作，具体如图 3 所示。

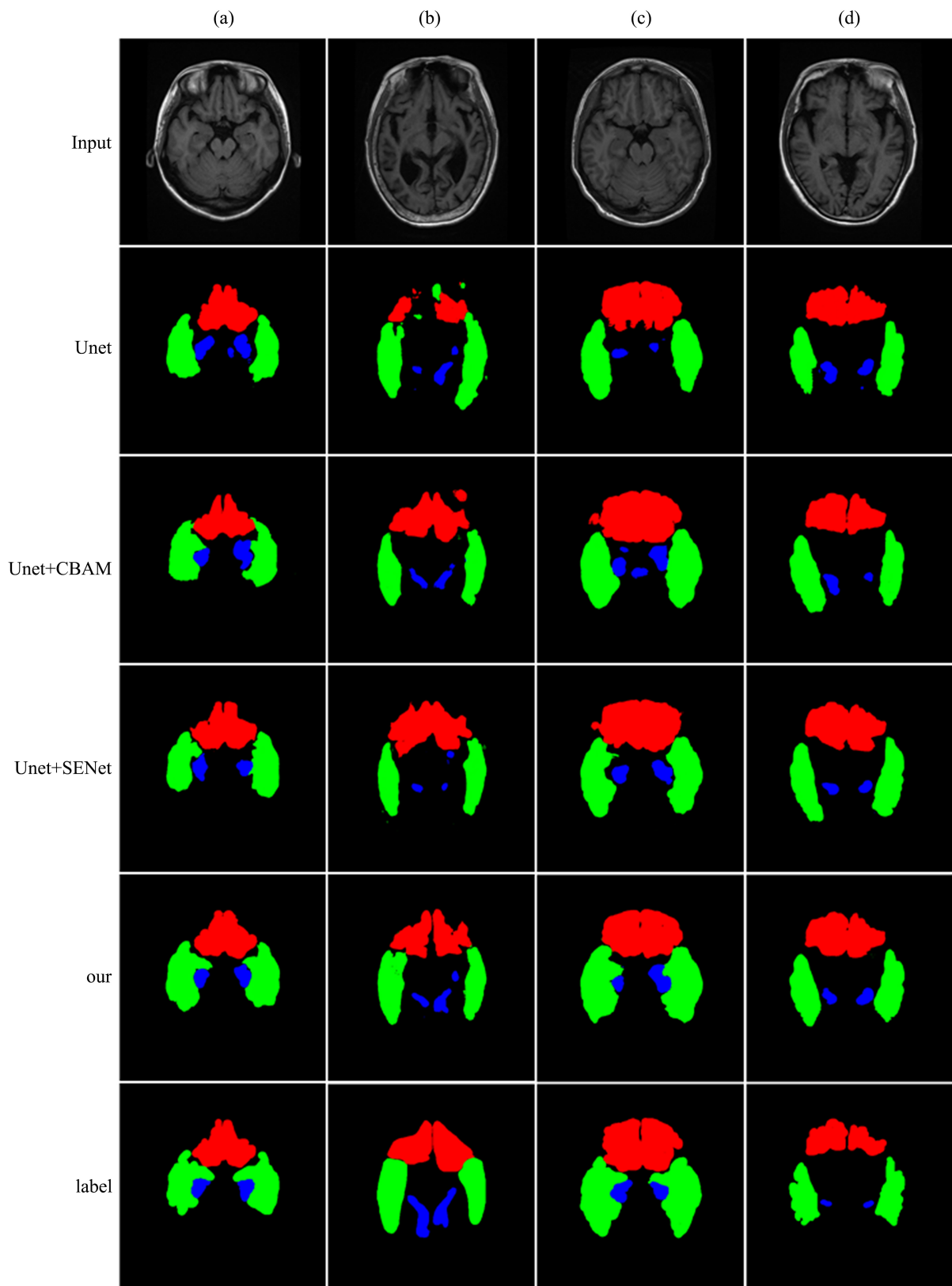


Figure 3. Comparison of experimental results: red (frontal lobe), green (temporal lobe), blue (hippocampus)
图 3. 实验结果对比: 红色(额叶)、绿色(颞叶)、蓝色(海马体)

5. 结论

本文将 CBAM 应用到了 Unet 的特征提取中, 并将 SENet 应用到了 Unet 的 Skip-connection 过程中。从而增强了分割部位的语义信息, 并降低了其它噪声对分割所带来的影响。同时, 在上采样的过程中模型可以更好的从全局中把握通道上的语义特征。实验表明, 本文中对原有的 Unet 模型进行的改进是有效的。

但是, 我们的模型还是存在着一些不足, 比如模型对小目标(海马体)分割的效果并不是很理想, 而对其他两个部位的分割的效果也还有一定的提升空间。而我们接下来的工作将主要针对模型整体分割效果以及小目标分割效果的提升来展开。由于医学图像标注需要一定的专业性, 所以带标签的医学图像数据集较少, 但是又有很多没有标签的医学图像。为了充分利用这些没有标签的图像, 我们下一步的工作也将会倾向于使用半监督或者无监督学习的方法。

基金项目

广东省重点领域研发计划项目(2019B010109001); 广东省信息物理融合重点实验室(2016B030301008); 广东省农业科学与技术创新团队项目(2019KJ147)。

参考文献

- [1] Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., *et al.* (2020) Convolutional Neural Networks for Classification of Alzheimer's Disease: Overview and Reproducible Evaluation. *Medical Image Analysis*, **63**, Article ID: 101694. <https://doi.org/10.1016/j.media.2020.101694>
- [2] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [3] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [4] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., *et al.* (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 424-432. https://doi.org/10.1007/978-3-319-46723-8_49
- [5] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., *et al.* (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [6] Huang, H., Lin, L., Tong, R., *et al.* (2020) Unet 3+: A Full-Scale Connected Unet for Medical Image Segmentation. *ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, 4-8 May 2020, 1055-1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>
- [7] Isensee, F., Jäger, P.F., Full, P.M., *et al.* (2020) nnU-net for Brain Tumor Segmentation. In: *International MICCAI Brainlesion Workshop*, Springer, Cham, 118-132. https://doi.org/10.1007/978-3-030-72087-2_11
- [8] Li, X., Chen, H., Qi, X., *et al.* (2018) H-Dense UNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, **37**, 2663-2674. <https://doi.org/10.1109/TMI.2018.2845918>
- [9] Alom, M.Z., Yakopcic, C., Hasan, M., *et al.* (2019) Recurrent Residual U-Net for Medical Image Segmentation. *Journal of Medical Imaging*, **6**, Article ID: 014006. <https://doi.org/10.1117/1.JMI.6.1.014006>
- [10] Wang, F., Jiang, M., Qian, C., *et al.* (2017) Residual Attention Network for Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 3156-3164. <https://doi.org/10.1109/CVPR.2017.683>
- [11] Nie, D., Gao, Y., Wang, L., *et al.* (2018) Asdnet: Attention Based Semi-Supervised Deep Networks for Medical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 370-378. https://doi.org/10.1007/978-3-030-00937-3_43
- [12] Sinha, A. and Dolz, J. (2019) Multi-Scale Guided Attention for Medical Image Segmentation.

- [13] Khosravan, N., Mortazi, A., Wallace, M., *et al.* (2019) Pan: Projective Adversarial Network for Medical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 68-76. https://doi.org/10.1007/978-3-030-32226-7_8
- [14] Son, J., Park, S.J. and Jung, K.H. (2017) Retinal Vessel Segmentation in Fundoscopic Images with Generative Adversarial Networks.
- [15] Xue, Y., Xu, T., Zhang, H., *et al.* (2018) SegAN: Adversarial Network with Multi-Scale L1 Loss for Medical Image Segmentation. *Neuroinformatics*, **16**, 383-392. <https://doi.org/10.1007/s12021-018-9377-x>
- [16] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [17] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>