[文章编号] 1003－4684(2025)01-0049-06

基于改进 U-net 的医学图像分割模型

杨永谱[1]，甘海涛[1]，夏 薇[2]，杨 智[1]，叶志伟[1]

(1 湖北工业大学计算机学院，湖北 武汉 430068；
2 华中科技大学同济医学院附属武汉儿童医院(武汉市妇幼保健院)，湖北 武汉 430070)

[摘 要] U-net 模型在医学图像分割中取得了很多的发展，但是其忽略了低层视觉特征和高层语义特征之间的差距，限制了网络的特征提取能力。针对这个问题，首先设计了一个特征增强模块，以更大的感受野来增强网络的特征提取能力，并使用跳跃连接将网络中低级特征与高级特征融合。同时引入三重注意力机制来融合空间维度和通道维度的特征，从而达到增强有用信息的目的。最后，设计了一种新的损失函数，引入二元交叉熵和 Dice 损失函数，来抑制数据集中类别不平衡的问题。该模型在 BraTs 数据集上进行了实验，取得了 0.8592 的平均 Dice。实验结果表明，该模型性能优于其他常用的分割网络。

[关键词] 医学图像分割；特征增强；注意力机制；损失函数
[中图分类号] TP391 [文献标识码] A

近年来，随着深度学习的快速发展，许多研究开始将深度学习应用到医学图像分割领域，并取得了巨大的成功[1]，展现了卷积神经网络(CNN)在医学图像分割领域的优势[2－3]。CNN 将医学图像作为输入，中间卷积层和池化层负责提取医学图像的特征，整个过程不需要人为干预，避免了传统机器学习中繁杂的操作[4－5]。随着研究的深入，有学者开发出了完全卷积神经网络(FCN)[6]，它以惊人的分割精度迅速成为了最流行的分割方法之一。后来，Ronneberger[7]等提出了一种名为 U-net 的全卷积神经网络用于医学图像分割并取得了巨大的成功。U-net 由编码器和解码器构成，在网络的内部还使用跳跃连接来融合不同层的特征，极大的提高了网络的特征提取能力。然而，在 U-net 网络及其各种改进的网络中都是使用相同尺寸的卷积块来提取特征，这在一定程度上限制了网络的特征提取能力。为解决这个问题，Zhang 等[8]在 U-Net 中引入 InceptionResNet 和密集连接，形成了一种新的语义分割网络，在肺、血管和脑瘤医学图像数据集上取得了较好的分割性能。但该网络同样具有一定的局限性，它忽略了不同特征之间的语义差距，在编码过程中容易丢失微小特征的信息。在此基础上，Yu 等[9]提出了一种深层聚合网络(DLANet)来解决编码过程中信息丢失的问题。该网络以深层聚合的方式来提取特征。在他们的语义分割任务中，深层聚合的思想是将两个模块迭代使用，从而形成一个层次逐渐加深、分辨率逐渐提高的解码器。U-net＋＋[10]也使用了深度聚合思想，但是与 DLANet 不同的是，U-net＋＋不仅采用迭代深度聚合策略，还增加了密集连接，从更深层来聚合语义特征，形成高度密集的特征聚合模型，达到减少信息丢失的目的。同时，Hu 等[11]在神经网络中引入注意力机制，通过注意力机制来放大有用信息抑制无用信息，从而提升网络分割性能。

综上所述，为了解决基于 U-net 的医学图像分割网络中卷积核尺寸单一且容易忽略不同特征之间语义关系和丢失微小特征信息的问题，本文提出了一种带三重注意力机制的特征增强分割模型(EFTA-Unet)，该模型能够充分利用医学图像中微小特征信息和捕获不同特征之间的语义关系并且能够克服医学图像数据集中类别不平衡的问题。

1 基于改进 U-net 的医学图像分割模型

本文提出了具有特征增强和注意力机制的医学图像分割模型(U-net with enhance feature and

[收稿日期] 2022－09－29
[基金项目] 湖北工业大学高层次人才基金(GCRC2020016)；湖北省自然科学基金(2021CFB282)；省部共建生物催化与酶工程国家重点实验室开放基金项目(SKLBEE2021020 和 SKLBEE2020020)
[第一作者] 杨永谱(1996－)，男，湖北黄冈人，湖北工业大学硕士研究生，研究方向为计算机视觉。
[通信作者] 甘海涛(1985－)，男，江西丰城人，湖北工业大学教授，研究方向为机器学习、计算机视觉和医学影像处理。

triplet attention module，EFTA-Unet)，其框架如图 1 所示。本文所提出的模型由编码器和解码器两部分组成。在编码器中通过级联的卷积层以及池化层来提取医学图像的语义特征。解码器则将这些特征恢复成像素级别的分割掩码。此外，EFTA-Unet 使用跳跃连接将编码器中的特征与解码器中的特征融合，减少编码过程中信息丢失。

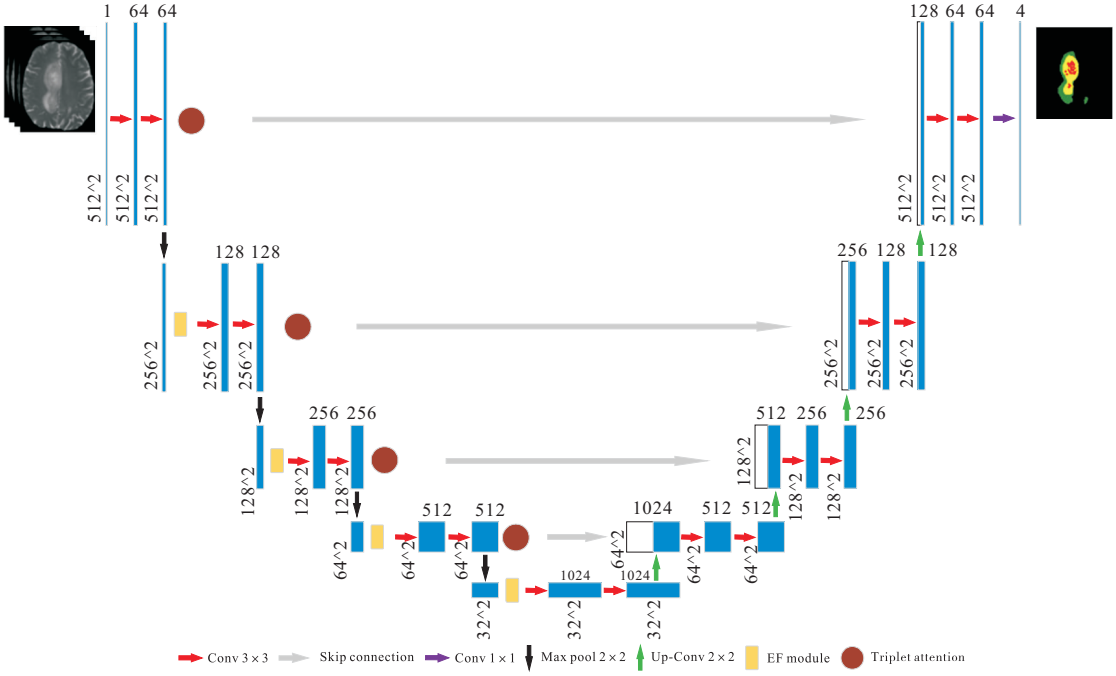


图 1 EFTA-Unet 结构

1.1 特征增强模块

图 2 是特征增强(enhance feature，EF)模块的结构。当特征输入时，会进行五个并联操作，在前面四个并联分支中，由两个卷积层组成。第一层的卷积核尺寸为 $N\times1$，第二层的卷积核尺寸为 $1\times N$。本文使用这种级联的非对称卷积，相对于 $N\times N$ 的对称卷积，能够以较小的计算量捕获同样多的特征。第五个分支是一个跳跃连接。最后将这五个分支的输出相加得到一个新的特征矩阵，新的特征矩阵经过卷积核大小为 3×3 和 1×1 的卷积层，最终得到整个分支的总输出。EF 模块具有不同尺寸的感受野，使网络具有更强的特征提取能力。

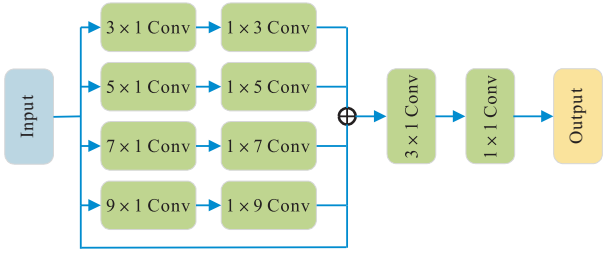


图 2 EF 模块结构

1.2 三重注意力机制(triplet attention，TA)

在通道注意力机制中，会对不同特征的权重进行均匀缩放。然而，这种方法需要通过全局平均池化将特征矩阵在空间上分解为单个像素来确定这些通道的权重。但是，由于它没有考虑到通道和空间之间的相互关系，所以会导致大量的空间信息丢失。为了解决这个问题，三重注意力机制中提出跨维度交互的技术，通过建模空间维度和通道维度之间的相关性来更新权重，达到缓解空间信息的丢失的目的。在这里，三个分支分别计算 (C,H)、(C,W) 和 (H,W) 之间的相关性。图 3 为三重注意力机制(triplet attention，TA)的结构，它具有三个并行分支。

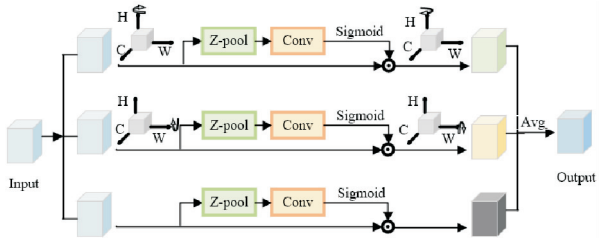


图 3 Triplet Attention 结构

如图 3 所示，三重注意力机制的第一个分支计算通道维度 C 和空间维度 W 相关性，第二个分支计算通道 C 和高度 H 之间的相关性，最后一个分支计算高度 H 和宽度 W 之间的相关性，即空间注意力。三个分支输出的特征矩阵尺寸相同。最后将三个分支的输出相加取平均值作为整个模块的输出。在三层注意力机制中，Z-pool 层的目的是减小特征维度数，Z-pool 操作能够在保留特征矩阵信息量的前提下降低特征矩阵的维度。Z-pool 的数学表达式：

$$Z\text{-}pool(X)=[MaxPool_{0d}(T),AvgPool_{0d}(T)] \quad (1)$$

其中，$T\in R^{C\times H\times W}$ 表示卷积层的输出，C 表示通道数，H 表示张量的高度，W 表示张量的宽度。$\mathrm{MaxPool}_{0d}(T)$ 表示在第 0 维执行最大池化操作，$\mathrm{AvgPool}_{0d}(T)$] 表示在第 0 维执行平均池化操作。对于尺寸为（$C\times H\times W$）的张量，经过 Z-pool 操作后张量的尺寸变为（$2\times H\times W$）。

在第一个分支中，对于一个输入特征 $T\in R^{C\times H\times W}$，进入第一个分支后，首先将特征矩阵沿 H 轴逆时针方向旋转 90°（$\hat{T}_1^*$）计算通道维度 C 与（$H\times W$）之间的依赖关系，得到尺寸为（$W\times H\times C$）的特征矩阵。将特征矩阵经过 Z-pool 操作，得到尺寸为（$2\times H\times C$）的特征矩阵。再将这个特征矩阵经过一个标准化卷积层和一个归一化操作，最终得到一个尺寸为（$1\times H\times C$）的特征矩阵。对这个输出执行 Sigmoid 激活函数，获得注意力权重。最后为了保证输出矩阵与输入矩阵尺寸一致，将得到的注意力权重矩阵沿着 H 轴顺时针方向旋转 90°（$\hat{T}_1$）。

第二个分支的作用是建模通道维度 C 和宽度维度 W 之间的相关性。首先将特征矩阵沿 W 轴逆时针旋转 90°（$\hat{T}_2^*$）得到尺寸为（$H\times C\times W$）的特征矩阵，再将这个特征矩阵经过 Z-pool 操作得到尺寸为（$2\times C\times W$）的特征矩阵并将这个特征矩阵经过一个标准卷积层和归一化操作，得到形状为（$1\times C\times W$）的注意力权重矩阵。最后，将得到的注意力权重经过 Sigmoid 操作和旋转操作，最终得到第二个分支的注意力权重矩阵，并沿着 W 轴顺时针方向旋转 90（$\hat{T}_2$）。

如图 3 所示，第三个分支与前两个分支不同。这个分支不执行旋转操作，对于输入的特征矩阵直接执行 Z-pool 操作，这个分支最终输出的特征矩阵形状为（$1\times H\times W$）。最后，将三个分支的输出融合成尺寸为（$C\times H\times W$）的输出作为整个注意力模块的输出。对于输入的特征矩阵 $\boldsymbol{T}\in R^{C\times H\times W}$，当经过三重注意力机制后得到的注意力矩阵 $\boldsymbol{S}$ 的过程如下：

$$\boldsymbol{S}=\frac{1}{3}\{\overline{\hat{T}_1\sigma(\varphi_1(\hat{T}_1^*))}+\overline{\hat{T}_2\sigma(\varphi_2(\hat{T}_2^*))}+T_3\sigma(\varphi_3(\hat{T}_3))\} \quad (2)$$

其中 σ 表示 sigmoid 激活函数；φ_1，φ_2，φ_3 表示标准的二维卷积层，T_3 和 $\hat{T}_3$ 等于 T。式(2)可进一步简化为：

$$\boldsymbol{S}=\frac{1}{3}(\overline{\hat{T}_1\omega_1}+\overline{\hat{T}_2\omega_2}+T_3\omega_3)=\frac{1}{3}(\bar{S}_1+\bar{S}_2+S_3) \quad (3)$$

其中 ω_1、ω_2 和 ω_3 分别表示三个分支的注意力矩阵。式(3)中的 $\bar{S}_1$ 和 $\bar{S}_2$ 表示顺时针和逆时针旋转操作。

1.3　损失函数

图像分割实际上是像素级别的分类任务，只需要将每个像素进行分类即可完成分割任务。对于脑胶质瘤图像分割任务而言，由于整个图像中除了肿瘤区域以外，其他的区域都可以算作是背景。所以二元交叉熵（binary cross entropy，BCE）可以作为损失函数的一部分，如

$$L_{\mathrm{BCE}}=\frac{1}{n}\sum_{i=1}^{n}(y_i\log\hat{y}_i+(1-y_i)\log(1-\hat{y}_i)) \quad (4)$$

式中：y_i 表示数据中真实的标签值，$\hat{y}_i$ 表示分割模型的预测值，n 为模型训练的批次数。由于脑胶质瘤 MRI 中肿瘤区域只占整个图像的 2%左右，背景占比太大，存在像素类别不平衡的问题。如果只用二元交叉熵作为损失函数，则会导致网络中在训练过程中偏向学习背景特征而疏于学习肿瘤区域的特征，进而影响网络的整体分割效果。而对于缓解类别不平衡问题最常用的损失函数为 Dice，所以也可以将 Dice 作为本文的损失函数的一部分，其公式如：

$$L_{\mathrm{Dice}}=1-\frac{2\sum_i y_i\hat{y}_i+\varepsilon}{\sum_i y_i+\sum_i\hat{y}_i+\varepsilon} \quad (5)$$

式中：y_i 表示真实标签值，$\hat{y}_i$ 表示模型预测的值，ε 是一个平滑算子，防止出现分子分母为 0 的情况。考虑到以上两个方面，本文结合这两个损失函数组成一个新的损失函数，公式如：

$$\mathrm{Loss}=\lambda L_{\mathrm{BCE}}+L_{\mathrm{Dice}} \quad (6)$$

其中 λ 为超参数。

2　实验设置与结果

2.1　数据集

为了验证 EFTA-Unet 模型，本文在公共数据集 BraTs 上进行了多次实验，并对实验结果进行评估。使用 BraTs2018 的训练集作为本文的训练集，共 285 例样本（HGG 为 210 例，LGG 为 75 例）。测试集则采用 BraTs2019 的训练集在 BraTs2018 基础上新增的 50 例样本（HGG 为 49 例，LGG 为 1 例）。每一例样本都包含四种模态，分别为 Flair、T1、T1ce 和 T2。MRI 的尺寸为 244×244×155，前两个数字尺寸表示图像的宽度和高度，第三个数字表示切片的数量。本文将三维的 MRI 沿轴位切片，分解为 155 张图片，然后丢弃黑色切片。需要注意的是，每位患者 MRI 切片中黑色图像的数量不一致，所以每位患者脑 MRI 的有用切片数量也不一

致。因此，对所有患者脑 MRI 处理后，用于实验的训练集图像有 18923 张，测试集的图像有 3138 张。用于实验的图像如图 4 所示(绿色代表水肿，黄色代表增强肿瘤，红色代表非增强肿瘤)。

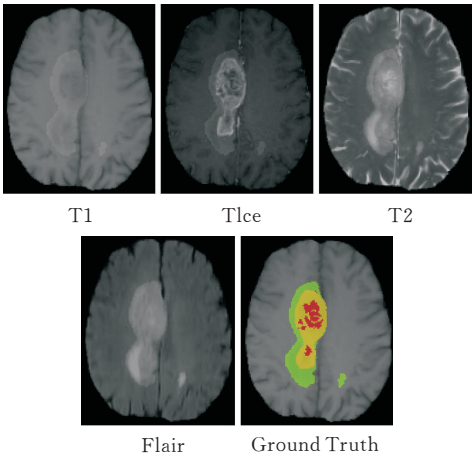


图 4 BraTs 数据可视化示例

脑胶质瘤的主要成分包括水肿(peritumoral edema，ED)、增强肿瘤(enhancing tumor，ET)、非增强肿瘤(non-enhancing tumor，NET)。本文的分割任务主要为整个肿瘤(WT，ED＋ET＋NET)，增强肿瘤(ET)，和核心肿瘤(TC，ET＋NET)。三个分割任务中，核心肿瘤(TC)和增强肿瘤(ET)相对于整个肿瘤(WT)而言属于小尺度肿瘤。

2.2 评价指标

本文使用 Dice 相关系数和豪斯多夫距离(Hausdorff95 distance)来衡量模型的性能。Dice 相关系数计算公式如：

$$\text{Dice}=\frac{2\text{TP}}{\text{FN}+\text{FP}+2\text{TP}} \tag{7}$$

其中，TP 为真阳性样本数，FP 为假阳性样本数，FN 为假阴性样本数。

豪斯多夫距离(Hausdorff95 distance)测量的是真实区域与预测区域之间的距离，它对分割区域的边界更为敏感，其计算公式如：

$$Haus95(T,P)=\max\{\sup_{t\in T}\inf_{p\in P}d(t,p),\ \sup_{p\in P}\inf_{t\in T}d(t,p)\} \tag{8}$$

其中，sup 表示最大值，inf 表示最小值，T 表示真实的肿瘤区域(Ground Truth)，P 表示预测的肿瘤区域，$d(t,p)$ 为距离函数，计算点 t 到 p 之间的距离。

2.3 实验细节

所有实验均在 Intel(R) Core(TM) i910940X CPU @ 3.30GHz，64GB RAM，NVIDIA GeForce RTX 3090 上进行。所有模型的优化器均为 Adam，学习率为 0.0001，批量大小为 8，本文的模型训练了 10000 个 epoch，损失函数中 λ 为 0.5。

2.4 消融实验

为验证本文所提网络(EFTA-Unet)的性能，进行了一系列的消融实验。模型的基础网络为 U-net，首先考虑只增加 EF 模块，记为 U-net＋EF，表 1 为对比实验结果。从表 1 中可以看出，引入 EF 模块后平均 Dice 值提升了 1.3%，性能更优。

表 1 消融实验，验证 EF 模块的有效性

Method	Dice			Average
	WT	TC	ET	
U-net	0.8387	0.8215	0.7688	0.8097
Unet＋EF	0.8472	0.8357	0.7846	0.8225
Method	Haus95			Average
	WT	TC	ET	
U-net	2.6170	1.7425	2.8341	2.3979
Unet＋EF	2.5816	1.6905	2.7563	2.3428

在前面实验的基础上，继续验证 TA 模块的有效性，名称标记方式与前面相似，实验结果如表 2 所示。从表 2 中可以看出，随着模块的增加，分割精度也在逐步提升，EFTA-Unet 表现出了最佳的分割性能。EFTA-Unet 的平均 Dice 值提升了 1.7%，性能更优。

表 2 消融实验，验证 AC 模块的有效性

Method	Dice			Average
	WT	TC	ET	
U-net	0.8387	0.8215	0.7688	0.8097
Unet＋AC	0.8512	0.8435	0.7842	0.8263
EFTA-Unet	0.8836	0.8627	0.8314	0.8592
Method	Haus95			Average
	WT	TC	ET	
U-net	2.6170	1.7425	2.8341	2.3979
Unet＋AC	2.5633	1.6883	2.7852	2.3456
EFTA-Unet	2.4716	1.6035	2.6743	2.2498

为了验证本文损失函数的有效性，进行了一系列的对比试验。首先本文单独使用 Dice 和二元交叉熵作为损失函数，记为 EFTA-Unet＋Dice 和 EFTA-Unet＋BCE，然后使用这两个损失函数的线性组合作为对比试验的损失函数，记为 EFTA-Unet＋Loss。实验结果如表 3 所示，使用本文所提损失函数的模型性能明显优于单独使用两个损失函数时的性能。

表 3 消融实验，验证损失函数 Dice、BCE 和 Loss 的有效性

Method	Dice			Average
	WT	TC	ET	
EFTA-Unet＋BCE	0.8415	0.8326	0.8124	0.8288
EFTA-Unet＋Dice	0.8357	0.8262	0.8092	0.8237
EFTA-Unet	0.8836	0.8627	0.8314	0.8592

2.5 实验结果

本文将 EFTA-Unet 与 U-net[7]，DeepResU-

net[14]，U-net ＋＋[10]，Dense _ Unet[12]，Unet3plus[13]在 BraTs 数据集上进行实验，实验结果如表 4 所示，图 5 为分割结果可视图。从表 4 中可以看出，EFTA-Unet 模型在 BraTs 数据集上取得了 0.8592的平均 Dice 和 2.2498 的平均 HD，EFTA-Unet 模型的分割性能明显优于常用的模型。

表 4　对比实验，与常用的分割模型在 BraTs 数据集上的分割结果对比

Method	Dice			Average
	WT	TC	ET	
U-net	0.8387	0.8215	0.7688	0.8097
DeepResUnet	0.8463	0.8319	0.7819	0.8200
Unet＋＋	0.8465	0.8604	0.7771	0.8280
Dense_Unet	0.8393	0.8703	0.7716	0.8271
Unet3plus	0.8183	0.7645	0.7368	0.7731

续表 4　对比实验，与常用的分割模型在 BraTs 数据集上的分割结果对比

Method	Haus95			Average
	WT	TC	ET	
EFTA-Unet	0.8836	0.8627	0.8314	0.8592
U-net	2.6170	1.7425	2.8341	2.3979
DeepResUnet	2.6811	1.7380	2.8103	2.4098
Unet＋＋	2.6530	1.6123	2.8291	2.3648
Dense_Unet	2.6421	1.5584	2.7939	2.3315
Unet3plus	2.7981	1.8893	2.9705	2.5526
EFTA-Unet	2.4716	1.6035	2.6743	2.2498

从图 5 中也可以看出，本文提出的模型在处理细节方面明显优于其他网络。整体来看，本文所提模型的分割性能最优。

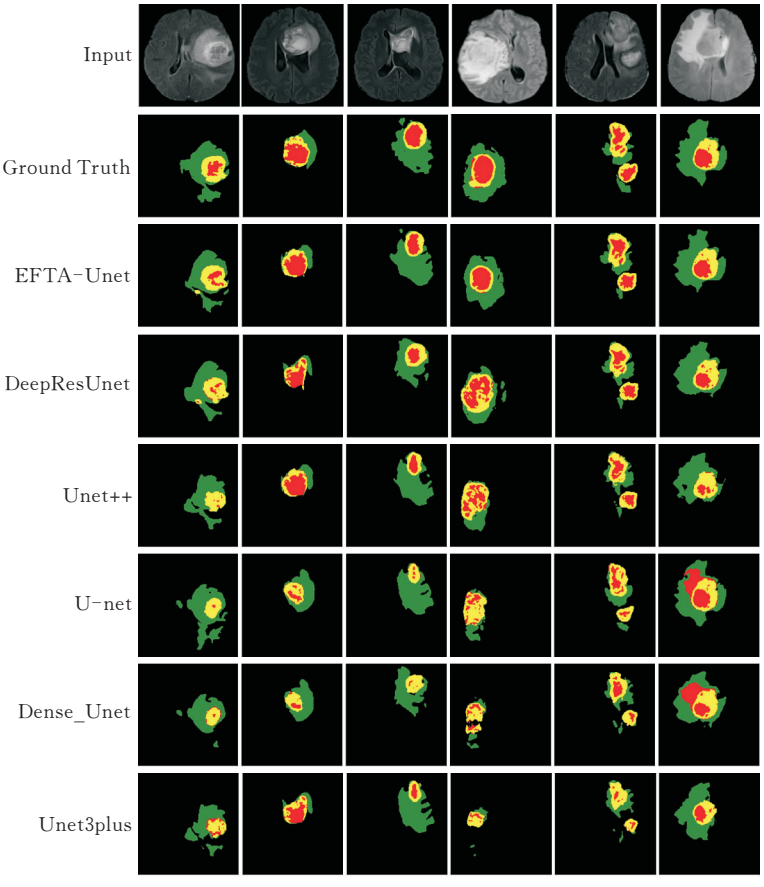


图 5　分割结果

3　总结

本文提出了一种名为 EFTA-Unet 的模型用于医学图像分割，该模型能够学习医学图像中微小特征信息和建模不同特征间的相关性，并且能够克服医学图像数据集中类别不平衡问题。本文提出的特征增强模块（EF 模块）具有不同尺寸的卷积方式，以不同大小的感受野提取特征，在一定程度上防止微小特征信息丢失。其次，引入三重注意力机制，它通过通道注意和空间注意来捕捉跨维度的特征并建立相关性。此外，本文设计了一种新的损失函数用来解决医学图像中类别不平衡问题。最后，将本文提出的 EFTA-Unet 模型与几种流行的医学图像分割模型在 BraTs 数据集上进行实验，可以发现与其他流行的分割模型相比，EFTA-Unet 模型在分割任务上表现最好。未来将进一步优化模型，希望在不影响分割性能的前提下减小模型的计算量。

[参 考 文 献]

[1] XIE X, NIU J, LIU X, et al. A survey on incorporating domain knowledge into deep learning for medical image analysis[J]. Medical Image Analysis, 2021, 69: 101985.

[2] WAN W, ZHONG Y, LI T, et al. Rethinking feature distribution for loss functions in image classification [C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 9117-9126.

[3] YANG H M, ZHANG X Y, YIN F, et al. Robust classification with convolutional prototype learning[C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 3474-3482.

[4] ZHENG Z, ZHONG Y, WANG J, et al. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery [C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 4096-4105.

[5] PATIL P W, BIRADAR K M, DUDHANE A, et al. An end-to-end edge aggregation network for moving object segmentation [C] // proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 8149-8158.

[6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 3431-3440.

[7] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[8] ZHANG Z, WU C, COLEMAN S, et al. DENSE-INception U-net for medical image segmentation [J]. Computer methods and programs in biomedicine, 2020, 192: 105395.

[9] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 2403-2412.

[10] XIAO X, LIAN S, LUO Z, et al. Weighted res-unet for high-quality retina vessel segmentation [C]. In: 2018 9th international conference on information technology in medicine and education (ITME), IEEE, 2018: 327-331.

[11] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(08): 2011-2023.

[12] PAN Y, LIU J, XU T, et al. Hippocampal segmentation in brain mri images using machine learning methods: A survey [J]. Chinese Journal of Electronics, 2021, 30(05): 793-814.

[13] HUANG H, LIN L, LONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation [C]. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020: 1055-1059.

[14] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual u-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(05): 749-753.

Medical Image Segmentation Model Based on Improved U-net

YANG Yongpu[1], GAN Haitao[1], XIA Wei[2], YANG Zhi[1], YE Zhiwei[1]

(1 *School of Computer Science, Hubei Univ. of Tech., Wuhan* 430068, *China*;
2 *Department Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong Univ. of Sci.& Tech., Wuhan* 430070, *China*)

Abstract: U-net model has made a lot of progress in medical image segmentation, but it ignores the gap between low-level visual features and high-level semantic features, which limits the feature extraction ability of the network. To solve this problem, a feature enhancement module is designed to enhance the feature extraction capability of the network with a larger receptive field, and a jump connection is used to fuse low-level features with high-level features in the network. At the same time, the triple attention mechanism is introduced to integrate the features of spatial dimension and channel dimension so as to enhance the useful information. Finally, a new loss function was designed, which introduced binary cross entropy and Dice loss function to suppress the category imbalance in the data set. The model was tested on the BraTs data set and the average Dice of 0.8592 was achieved. Experimental results show that the performance of this model is better than other common segmentation networks.

Keywords: medical image segmentation; feature enhancement; attention mechanism; loss function

[责任编校: 裴 琴]