

引用格式: ZHOU Tao, DONG Yali, LIU Shan, et al. Cross-modality Multi-encoder Hybrid Attention U-Net for Lung Tumors Images Segmentation[J]. Acta Photonica Sinica, 2022, 51(4):0410006

周涛,董雅丽,刘珊,等.用于肺部肿瘤图像分割的跨模态多编码混合注意力U-Net[J].光子学报,2022,51(4):0410006

用于肺部肿瘤图像分割的跨模态多编码混合 注意力U-Net

周涛^{1,2},董雅丽¹,刘珊¹,陆惠玲³,马宗军⁴,侯森宝¹,邱实⁵

(1 北方民族大学 计算机科学与工程学院,银川 750021)

(2 北方民族大学 图像图形智能处理国家民委重点实验室,银川 750021)

(3 宁夏医科大学 理学院,银川 750004)

(4 宁夏医科大学总医院骨科,银川 750004)

(5 中国科学院西安光学精密机械研究所,西安 710119)

摘要:针对医学图像的肺部肿瘤分割中病灶和周围组织的对比度低、边缘模糊、肿瘤和正常组织粘连、病灶和背景分布不均衡等问题,提出跨模态多编码混合注意力机制模型分割肺部病灶,用多种模态医学图像辅助分割病灶。首先设计了三编码器提取多模态医学图像的病灶特征,解决单模态医学影像的病灶特征提取能力不足的问题;然后针对网络通道维度冗余和对复杂病灶的空间感知能力不高的问题,在网络跳跃连接中加入混合注意力机制;最后对网络解码路径不同的尺度特征使用多尺度特征聚合块充分利用各个尺度特征。在临床多模态医学图像数据集上验证算法的有效性,对比实验结果表明所提模型对于肺部病灶分割的戴斯相似系数、召回率、体积重叠误差和相对体积差异分别为 96.4%、97.27%、93.0%、93.06%。对于病灶形状复杂,病灶和正常组织粘连的情况,分割精度得到有效提升。

关键词:深度学习;医学图像分割;多模态医学图像;U-Net;肺癌

中图分类号:TP391.41

文献标识码:A

doi:10.3788/gzxb20225104.0410006

0 引言

肺癌是一种常见且恶性程度高的癌症^[1],5年生存率仅为18%^[2],严重威胁着人们的健康。医学影像技术^[3]对肺部肿瘤的临床治疗有重要意义,其无创的成像方式被广泛应用在肿瘤诊断、分期、疗效评估等多个阶段,然而目前医学图像病灶分割仍存在一些挑战:1)病灶勾画依赖有经验的临床医生手动分割,人体复杂的解剖结构导致该工作费时费力;2)肺部肿瘤图像对比度低、病灶大小和形状不一、病灶位置多变,导致精准分割困难;3)肺部肿瘤图像数据类别分布不平衡,存在大量背景信息干扰。

医学图像分割方法分为传统方法和深度学习方法,传统分割方法^[4-6]依赖参数的设定,对初始轮廓敏感,对复杂形状和对比度低的图像分割效果不理想。以卷积神经网络(Convolutional Neural Networks, CNNs)为代表的深度学习方法在计算机视觉中迅速发展。医学图像分割要求分割结果更加准确,然而数据集少、缺乏标签等特点给医学图像分割带来了困难。2015年,RONNEBERGER O^[7]提出网络架构对称的U-Net,在少量医学数据的情况下仍能获得良好的分割结果,应用于多种医学图像分割场景^[8-9]。黄鸿^[10]提出基于U-Net++改进的肺结节分割算法,从网络不同深度提取语义信息,然后通过权重聚合模块自适应学习网络不同层的特征权重。ZHANG J^[11]提出LCU-Net(Low-cost U-Net)用于环境微生物图像分割,网络使用

基金项目:国家自然科学基金(No.62062003),宁夏自治区重点研发计划(No.2020BEB04022),北方民族大学引进人才科研启动项目(No.2020KYQD08),宁夏自然科学基金(No.2022AAC03149)

第一作者:周涛(1977-),男,教授,博士,主要研究方向为医学图像分析处理、计算机辅助诊断、模式识别。Email:zhoutaonxmu@126.com

通讯作者:董雅丽(1997-),女,硕士研究生,主要研究方向为医学图像分析处理、深度学习、模式识别。Email:dongyali_dyl@163.com

收稿日期:2021-10-18;**录用日期:**2021-12-30

<http://www.photon.ac.cn>

Inception 模块增大感受野,针对改进 U-Net 的分割结果使用密集条件随机场进行后处理。但是 U-Net 网络存在三个问题:1)U-Net 网络对每一幅特征图使用统一的参数,对于大小不一、形状复杂的病灶,网络可能出现空间感知不高,分割性能下降;2)U-Net 未充分考虑不同通道对特征的表达能力,影响模型的鲁棒性和分割性能;3)大多数多编码器 U-Net^[12-13]通过提取单模态医学图像目标切片的信息,提高网络分割性能,但是未充分利用不同模态医学图像对病灶的特征表达能力。

针对以上问题,本文提出基于多编码器混合注意力机制的 U-Net 网络(Multi-Encoder Attention U-Net, MEAU-Net),将病灶的多个模态医学图像同时输入到网络;使用混合注意力机制,其中通道注意力机制使得网络关注重要的特征通道,空间注意力机制增强网络对病灶部位的聚焦能力;最后使用多尺度特征聚合模块,对网络解码阶段得到尺度不同的特征图聚合,实现不同尺度特征的充分利用。

1 网络架构

基于 U-Net 提出多编码器混合注意力机制网络 MEAU-Net。网络包括提取多模态医学图像特征的编码器、还原病灶图像语义特征的解码器,在跳跃连接部分添加混合注意力机制,在解码路径使用多尺度特征聚合块。正电子发射断层扫描图像(Positron Emission Tomography, PET)图像提供病灶功能信息,计算机断层扫描成像(Computed Tomography, CT)图像提供病灶的解剖信息,正电子发射计算机断层显像(Positron Emission Tomography/Computed Tomography, PET/CT)结合两者的优点,多模态医学图像包含的病灶信息比单模态图像更丰富,因此网络同时向编码器中输入 PET/CT、PET 和 CT 的感兴趣区域(Region of Interest, ROI)图像提取三种模态图像病灶特征。在网络跳跃连接部分,使用混合注意力机制,包括空间注意力机制和通道注意力机制。受文献^[14]启发,对 PET/CT 和 CT 图像特征使用空间注意力机制,强调特征图中病灶区域,抑制无关背景。使用通道注意力机制对 PET/CT、CT 和 PET 三个分支提取对应通道的权重值,突出网络中特征图重要通道。针对解码路径得到精细程度不同的多尺度特征图,使用多尺度特征聚合块充分利用病灶不同尺度特征图,从而实现了对图像病灶的精准分割。

1.1 三编码器网络

MEAU-Net 编码路径包括 PET/CT、PET 和 CT 三个分支,其中 CT 和 PET/CT 分支由四个下采样块组成,PET 分支由三个下采样块组成。每一个下采样块包括两个卷积块和最大池化操作。卷积块包含 3×3 Same 卷积(指输入与输出特征图尺寸相同),批归一化(Batch Normalization, BN)和激活函数 ReLU,卷积步长和填充均为 1。通过上述两次卷积之后,使用 2×2 最大池化进行下采样。网络最后一层将 PET/CT 和 CT 分支的特征进行叠加,然后使用 1×1 卷积将维度降低到 1 024,将降维后的特征图传输到解码路径。解码路径由四个上采样块组成,每个上采样块由与编码路径一致的两个卷积块、一个 2×2 转置卷积组成。在跳跃连接部分,将经过混合注意力机制的特征图输入到解码路径。针对解码路径得到不同尺度特征,使用多尺度特征聚合块对跨尺度的特征进行聚合,最终输出分割结果。图 1 为 MEAU-Net 网络架构。

MEAU-Net 网络基于编码器和解码器两个部分,编码器用于捕获 PET/CT、PET 和 CT 图像的低级特征,每个编码块包括两次 3×3 卷积操作、BN 和 ReLU。过程表示为

$$\chi_i^{l+1} = \text{Conv}_{3 \times 3}(\chi_i^l), i = 1, 2, 3 \quad (1)$$

式中, χ_1^l 表示编码器中第 l 层的 PET/CT 特征图, χ_2^l 为第 l 层的 CT 特征图, χ_3^l 为第 l 层的 PET 特征图, $\text{Conv}_{3 \times 3}(\cdot)$ 表示对输入的三种特征图进行 3×3 卷积操作、BN 和 ReLU 操作。

解码器通过上采样和跳跃连接操作逐步恢复图像的空间信息,网络拼接编码器通过混合注意力机制的特征和解码器后一层的特征进行上采样。混合注意力机制使网络关注在图像病灶部分和重要通道,将编码器对应层的 PET/CT、PET、CT 三个分支的特征经过空间和通道注意力机制输入到解码器,参与网络的上采样操作。上采样操作表示为

$$\varphi^l = U[\varphi^{l-1} + A(\chi_1^l, \chi_2^l, \chi_3^l)] \quad (2)$$

式中, $U(\cdot)$ 表示两次卷积操作和转置卷积操作, φ^{l-1} 表示网络解码器 $l-1$ 层的特征图, $+$ 表示特征拼接, $A(\cdot)$ 表示混合注意力机制操作。编码器分别包含 PET、PET/CT 和 CT 三个分支的编码卷积块。解码器共四层,每一层包括上采样块和编码器对应层特征。

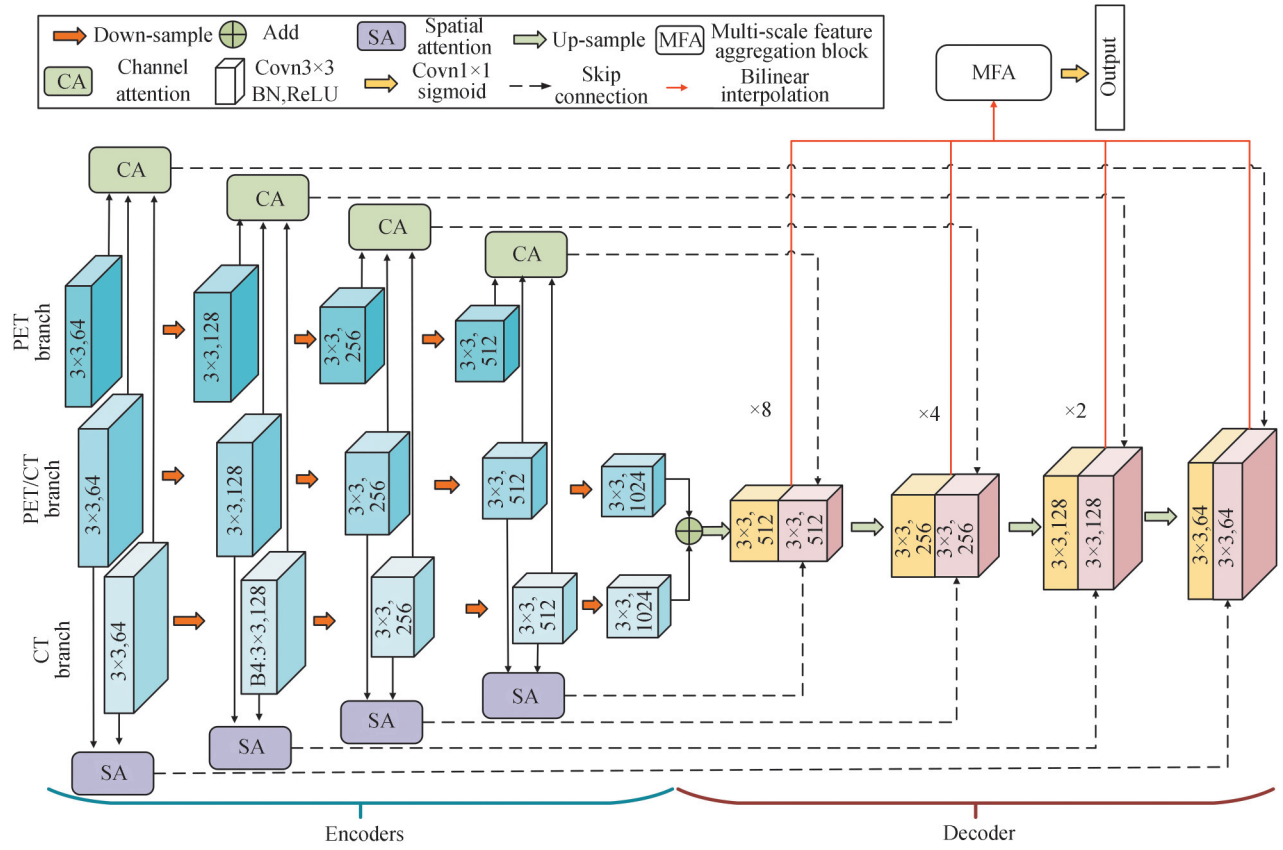


图1 MEAU-Net 网络架构
Fig. 1 MEAU-Net network architecture

1.2 混合注意力机制

网络跳跃连接引入空间注意力机制和通道注意力机制,对重要相关特征施加更多权重,允许网络专注于输入的特定部分,而不是输入每个特征。针对不同通道和空间特征对分割任务的不同贡献,对其分配不同的权重。

1.2.1 空间注意力机制

使用空间注意力机制聚焦特征图中病灶部分,抑制背景等无关信息。PET 成像技术对高代谢的组织 and 病灶呈现高亮信号,存在空间分辨率低的问题,不能提供准确的病灶边缘。所以空间注意力机制使用 PET/CT 和 CT 两种医学图像。算法伪代码(算法 1)表示为

Algorithm 1: Spatial attention mechanism

Inputs: The input map of the two branches features of PET/CT and CT $\chi_i^i, i = 1, 2$

Output: SAⁱ1:

- 1: $h_1^i = \text{concat}(\chi_1^i, \chi_2^i)$ /* Add the feature maps of PET/CT and CT */
- 2: $h_{\text{mean}} = \text{AvgPool}(h_1^i)$ /* avg-pooling */
- 3: $h_{\text{max}} = \text{MaxPool}(h_1^i)$ /* max-pooling */
- 4: $f = \text{concat}(h_{\text{mean}}, h_{\text{max}})$ /* Concatenate the feature map of avg-pooling and max-pooling */
- 5: $\beta = \text{Conv}_{3 \times 3}(f)$ /* 3×3 convolution operation */
- 6: $z = \sigma(\beta)$ /* After sigmoid, the feature map becomes $C \times H \times 1$ */
- 7: $\text{SA}^i = z \times h_1^i + h_1^i$ /* The feature map of sigmoid is multiplied with the original feature and then add */

End

首先对输入的两特征图进行叠加,分别对其平均池化和最大池化,平均池化对肺部病灶图像去噪,最大池化突出医学图像中病灶部分;然后拼接两种池化之后的特征图,对拼接后的特征图使用 3×3 卷积操作,

使用 sigmoid 将特征值压缩到 0 到 1 之间, 叠加的特征图与 sigmoid 后的权重值相乘, 最后相乘后的特征图与初始特征图叠加, 图 2 为空间注意力机制示意图。

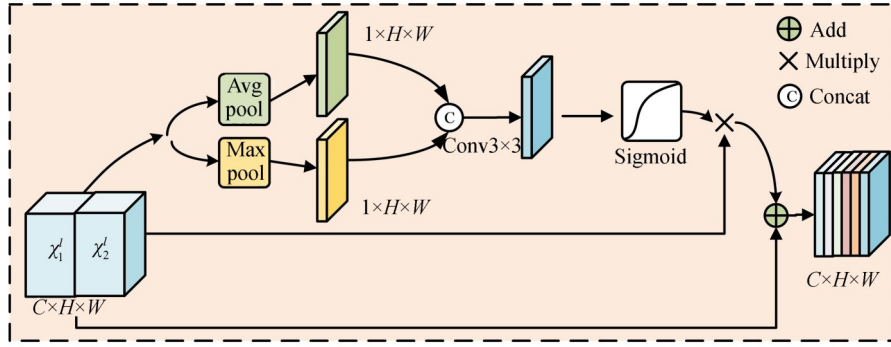


图 2 空间注意力机制
Fig. 2 Spatial attention mechanism

空间注意力机制具体表示为

$$SA^l(F) = (\chi_1^l + \chi_2^l) \oplus \{ (\chi_1^l + \chi_2^l) \times \{ \sigma \{ \text{Conv}_{3 \times 3} [\text{AvgPool}(\chi_1^l + \chi_2^l) + \text{MaxPool}(\chi_1^l + \chi_2^l)] \} \} \} \quad (3)$$

式中, F 表示操作的特征图, $SA(\cdot)$ 表示空间注意力机制操作, χ_1^l 和 χ_2^l 分别为当前 l 层的 PET/CT 和 CT 特征图, AvgPool 表示平均池化, MaxPool 表示最大池化, σ 表示 sigmoid 操作, \oplus 表示特征图叠加。

1.2.2 通道注意力机制

MEAU-Net 通过跳跃连接将编码路径特征拼接和解码路径, 对编码路径三种特征进行通道注意力很有必要。使用通道注意力机制对来自编码器的三种模态特征图进行重要通道加权, 对重要的通道分配更多的权重。通道注意力机制伪代码(算法 2)表示为

Algorithm 2: Channel attention mechanism

Inputs: The input features map of the three branches of PET/CT, CT and PET $\chi_i^l, i = 1, 2, 3$

Output: $CA^l(F)$

1: $\chi_{\text{hybrid}}^l = \chi_2^l \oplus \chi_3^l$ /* Add the feature maps of PET and CT, χ_{hybrid}^l */

2: $\chi^l = \chi_{\text{hybrid}}^l + \chi_1^l$ /* Concatenate the feature map of χ_{hybrid}^l and PET/CT */

3: $h^l, g^l = \text{AvgPool}(\chi^l), \text{MaxPool}(\chi^l)$ /* Avg-pooling and max-pooling on the feature maps respectively */

4: $\alpha = \sigma(\text{MLP}(h^l) + \text{MLP}(g^l))$ /* Perform MLP operations on h^l, g^l separately */

5: $CA^l(F) = \alpha \times \chi^l + \chi^l$

PET/CT、CT 和 PET 病灶图像提供的病灶信息不同, 由于 PET 图像反映病灶代谢信息, CT 图像包含丰富的病灶解剖信息, 首先将来自编码路径的 CT 特征图 χ_2^l 与 PET 特征图 χ_3^l 叠加, 将叠加之后的特征图 χ_{hybrid}^l 与 PET/CT 特征图 χ_1^l 拼接。为了保留特征图中更多信息, 分别使用平均池化和最大池化对拼接的特征图处理得到两个 $1 \times 1 \times C$ 权重值。接着使用由两个全连接层和 ReLU 组成多层感知机 (Multi-Layer Perceptron, MLP), 其中第一个全连接层有 $C/3$ 个神经元, 连接 ReLU, 第二个全连接层有 C 个神经元。最后将 MLP 得到的两个结果进行和操作, 经过 Sigmoid 得到 α , 将 α 与 χ^l 相乘之后再叠加。图 3 为通道注意力机制示意图。

通道注意力机制可表示为

$$CA^l(F) = \sigma \{ \text{MLP} \{ \text{AvgPool} [(\chi_2^l \oplus \chi_3^l) + \chi_1^l] \} + \text{MLP} \{ \text{MaxPool} [(\chi_2^l \oplus \chi_3^l) + \chi_1^l] \} \} \quad (4)$$

式中, $CA(\cdot)$ 表示通道注意力机制。最后将空间注意力机制的特征图与通道注意力的特征图相叠加, 即为混合注意力机制的特征图, 输入到对应层的解码器中。混合注意力可表示为

$$A^l(\chi_1^l, \chi_2^l, \chi_3^l) = SA^l(F) \oplus CA^l(F) \quad (5)$$

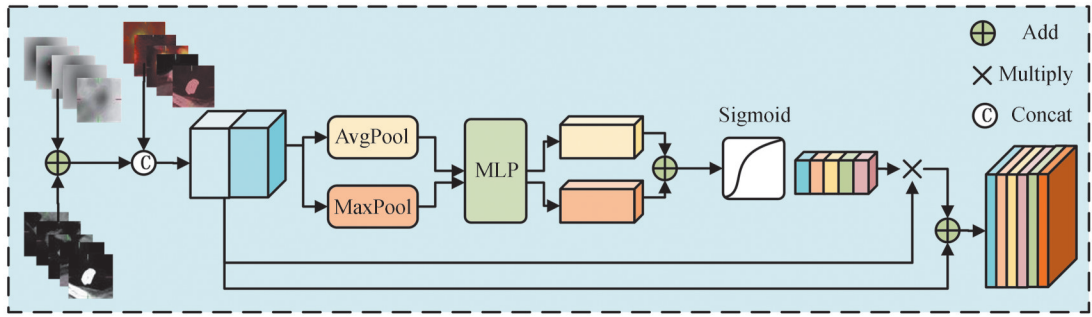


图3 通道注意力机制

Fig. 3 Channel attention mechanism

1.3 多尺度特征聚合块

网络解码阶段得到尺度不同的特征图,这些特征图包含编码路径的低级特征和解码路径的语义特征。由于不同尺度得到的特征图大小不同,通过上采样放大的特征图的像素相关性是不同的,直接对放大的特征图处理可能会丢失信息。因此,受文献^[15]启发,使用多尺度特征聚合(Multi-Scale Feature Aggregation, MFA)模块,模块示意图如图4所示。首先,使用双线性插值将不同尺度特征图放大到原图像 $50 \text{ pixel} \times 50 \text{ pixel}$ 大小,然后使用 1×1 卷积将四个尺度特征分别压缩为16,接着将其拼接,然后进行平均池化和MLP得到通道系数 α ,即

$$\alpha = \sigma \{ \text{MLP} [\text{AvgPool}(F)] \} \quad (6)$$

通道系数与拼接的特征图相乘之后进行 3×3 卷积、ReLU和 1×1 卷积、Sigmoid,得到系数 β ,即

$$\beta = \sigma \{ \text{Conv}_{1 \times 1} [\text{Conv}_{3 \times 3}, \text{ReLU}(F \cdot \alpha)] \} \quad (7)$$

最后使用残差连接将特征连接起来。具体过程表示为

$$F_{\text{MFA}} = F \times \alpha \times \beta + F \times \alpha + F \quad (8)$$

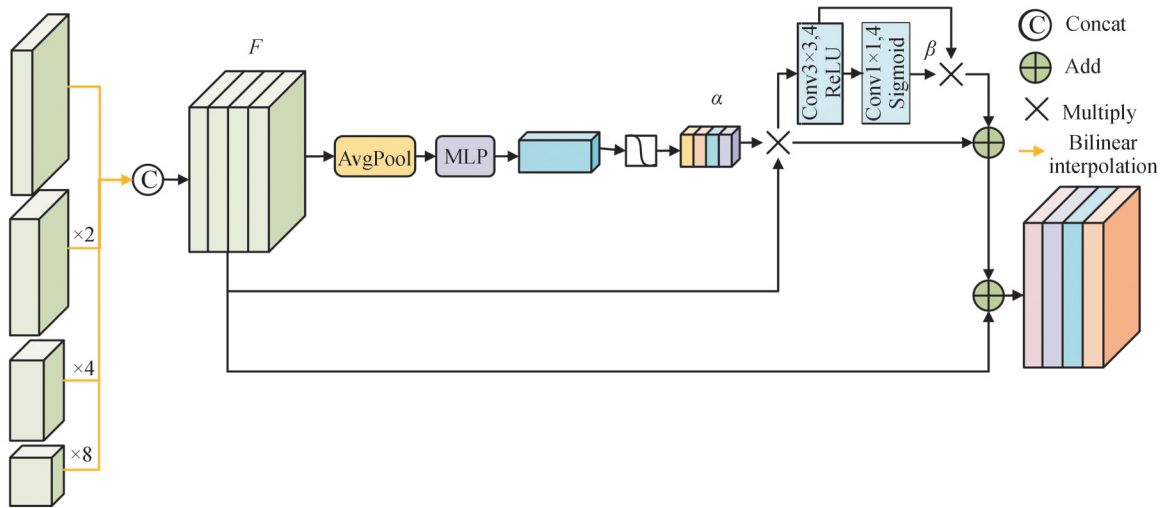


图4 多尺度特征聚合块

Fig. 4 Multi-scale feature aggregation block

2 实验结果与分析

2.1 数据集及数据预处理

CT是诊断肿瘤工具之一,它能够反映病灶的解剖信息。然而CT对于肿瘤微小病灶和原发灶的检出率低,容易导致漏诊。肿瘤组织相比正常组织生长迅速、代谢旺盛,PET利用这些特点将示踪剂注入人体内使得标记物在病灶内聚集。PET/CT将功能成像PET和解剖成像CT两种成像技术相结合,综合后的图像将病灶的解剖信息和功能信息反映在同一幅影像上,得到肿瘤的位置、形态、密度、代谢信息和大小。从而达

到“1+1>2”的效果,得到更为准确的病灶信息,实现对疾病的“定位”、“定性”、“定量”、“定期”。 ^{18}F FDG PET/CT对肺癌的治疗至关重要,其能无创地确定转移病灶和预测肿瘤特征^[16],广泛应用于肺部肿瘤的良恶性诊断^[17]、分期^[18]、预后评估^[19]和治疗方案制定^[20]的决策等。如图5所示CT图像中肉眼很难分辨病灶和正常组织的密度差异,而PET/CT图像中病灶代谢旺盛,呈高亮,因此多模态医学图像可以更好地定位病灶,识别病灶边缘。

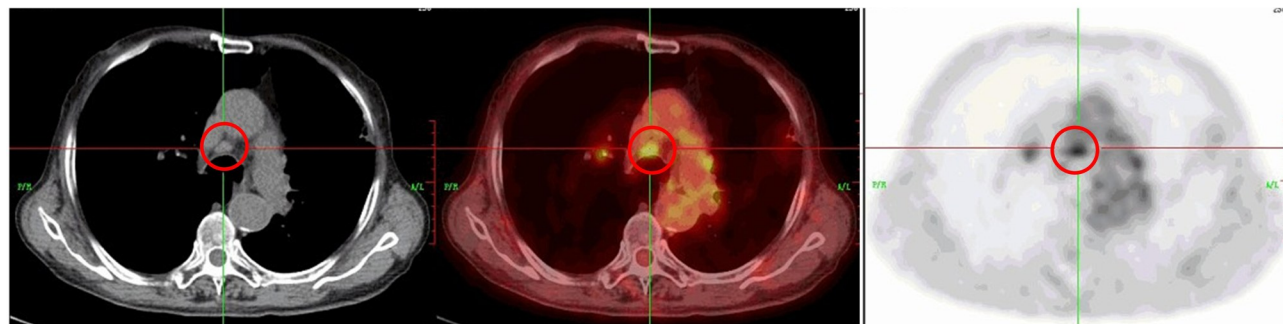


图5 CT、PET/CT和PET图像
Fig. 5 CT, PET/CT and PET image

选用2018年1月-2019年6月在宁夏某三甲医院核医学科进行PET/CT全身检查的90例肺部肿瘤临床患者,包括女性32例,男性58例。获取患者已配准的PET/CT、PET和CT二维肺部肿瘤图像,患者年龄范围为26~82岁,平均年龄为60岁。PET/CT、PET和CT图像各90张,该数据集标签均由临床医生参考肺部多模态医学图像后手动勾画。经过旋转和镜像等数据增广处理,最终三种模态图像数据集样本个数分别为1026张,其中909张划分为训练集,117张划分为测试集。肺部医学图像存在病灶和背景分布不平衡问题,网络容易学习到图像大量背景信息,导致分割结果不准确。肺部病灶由于其本身的成像特点给图像分割带来一定的困难,如人体解剖结构复杂,组织间相互重叠导致病灶边缘模糊^[21]。因此对获取的医生标注CT、PET和PET/CT二维切片尺寸为356 pixel×356 pixel,使用Hough变换^[22]对医生标注的肺部CT图像进行直线检测获取感兴趣区域,感兴趣区域大小为50 pixel×50 pixel。首先将彩色图像转化成灰度图像,使用5×5高斯滤波对CT图像进行模糊处理,然后使用Canny边缘检测算法获得图像中每个像素的边缘梯度和梯度方向,其次进行Hough变换,将图像从直角坐标系映射到Hough空间,将图像中的边缘点从直角坐标系转换到极坐标的参数方程,进而找到图像中相交的两条病灶的标注直线,最后对CT、PET和PET/CT图像进行50 pixel×50 pixel的裁剪,得到ROI图像。针对医学图像中病灶与背景对比度低的问题使用基于曝光融合^[23]的图像对比度增强方法来提高肺部病灶CT图像的对比度。

2.2 评价指标

为评估MEAU-Net性能,采用戴斯相似系数(Dice Similariy Coefficient, DSC)、召回率(Recall)、体积重叠误差(Volumetric Overlap Error, VOE)和相对体积差异(Relative Volume Difference, RVD)来评估模型分割性能。4种评价指标的公式定义分别为

$$\text{DSC} = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{VOE} = \text{abs} \left(1 - \left| \frac{P \cap G}{P \cup G} \right| \right) \quad (11)$$

$$\text{RVD} = \text{abs} \left(\frac{P}{G} - 1 \right) \quad (12)$$

式中,病灶区域被正确分割的定义为真阳性(True Positive, TP),正常组织区域被分割为病灶区域定义为假阳性(False Positive, FP),正常区域被正确分割定义为真阴性(True Negative, TN),病灶区域分割为正常区域

定义为假阳性(False Negative, FN), P 表示模型预测的目标像素, G 表示标签值(ground truth)中的目标像素。

2.3 实验环境与参数设置

实验硬件环境:服务器Intel XI(R) Gold 6154 CPU,内存256GB,显卡NVIDIA TITAN V,python3.7,PyTorch1.7.0,CUDA版本为11.1.106。使用Adam优化器,训练次数为200,学习率初始化为0.005,批处理大小为8,为了防止网络过拟合,在网络层中加入dropout,取值为0.5。损失函数是用来评估网络模型的预测值和真实值差异的函数,损失函数越小,意味着模型的预测值和真实值之间的差异越小。使用交叉熵损失函数,其中 y' 表示模型的输出,定义为

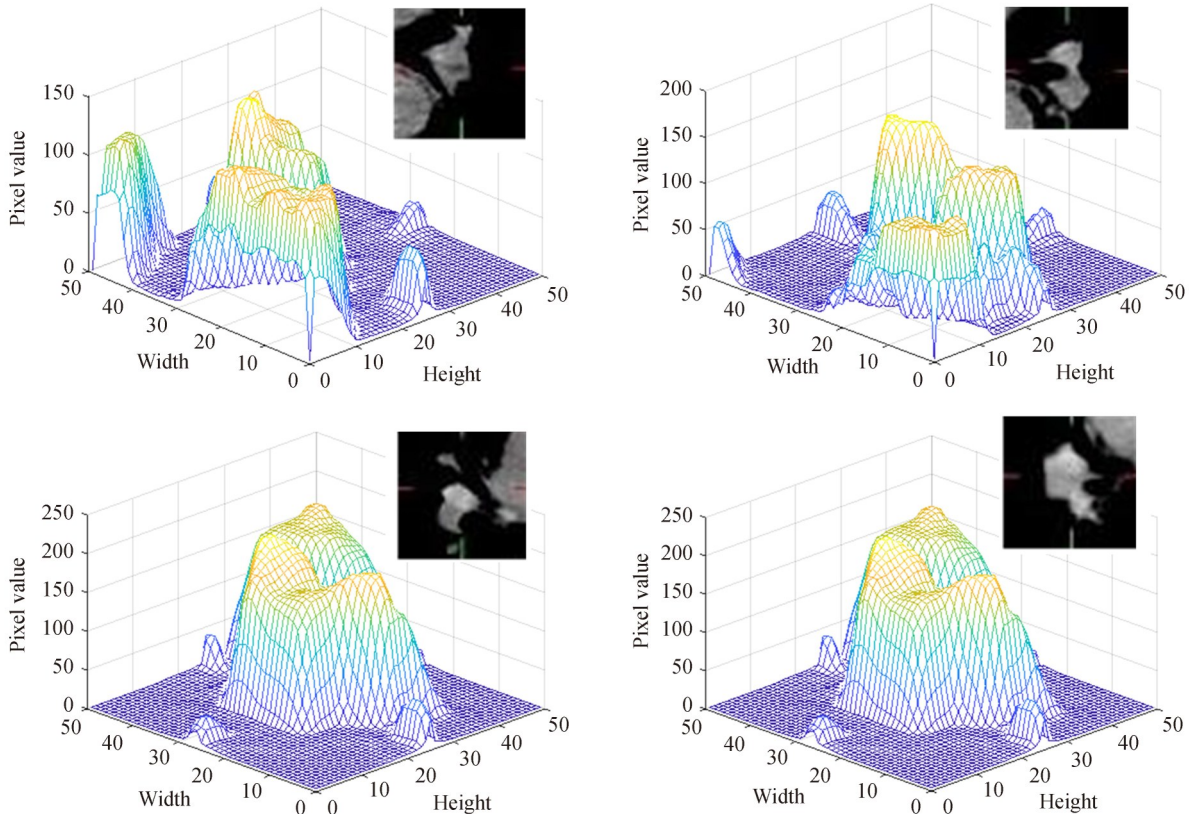
$$L = -y \log y' - (1 - y) \log (1 - y') \quad (13)$$

2.4 结果与分析

通过两组实验来说明MEAU-Net模块的先进性,第一组实验验证多编码器对网络分割的影响;第二组实验在三编码器U-Net基础上,探索混合注意力机制对模型性能的影响;最后,将MEAU-Net与先进分割方法作比较。需要说明的是,采用DSC、Recall、VOE和RVD等4个评价指标对算法进行评价,但是体积重叠误差VOE和相对体积差异RVD是越小越好,为了统一4个评价指标,VOE和RVD两个评价指标的值是取1与这两个指标的差值。

2.4.1 不同编码器分割网络

通过三个实验对比不同编码器对分割效果的影响。实验一使用原始U-Net^[8]仅输入单模态CT的图像,使用909张CT图像作训练集,117张CT图像作测试集。实验二使用两编码器的Y-Net^[24],该网络包括两个编码器和一个解码器,将同层的两个编码器提取的特征与解码器下层提取的特征叠加输入到对应层的解码器中,两个解码器分别输入CT和PET图像,使用PET和CT图像各909张作训练集,PET和CT图像各117张作测试集。实验三使用三编码器U-Net模型MEU-Net (Multi Encoder U-Net),在编码器最后一层将三个编码器提取的多模态医学图像低级特征叠加,由于PET图像的病灶轮廓模糊,将编码层PET/CT和CT特征图拼接后经跳跃连接输入对应解码层,使用PET/CT、CT和PET图像各909张作训练集,PET/CT、CT和PET图像各117张作测试集。CT图像的三维灰度图如图6所示,不同编码器网络分割结果如表1所示,图7为不同编码器架构分割结果。



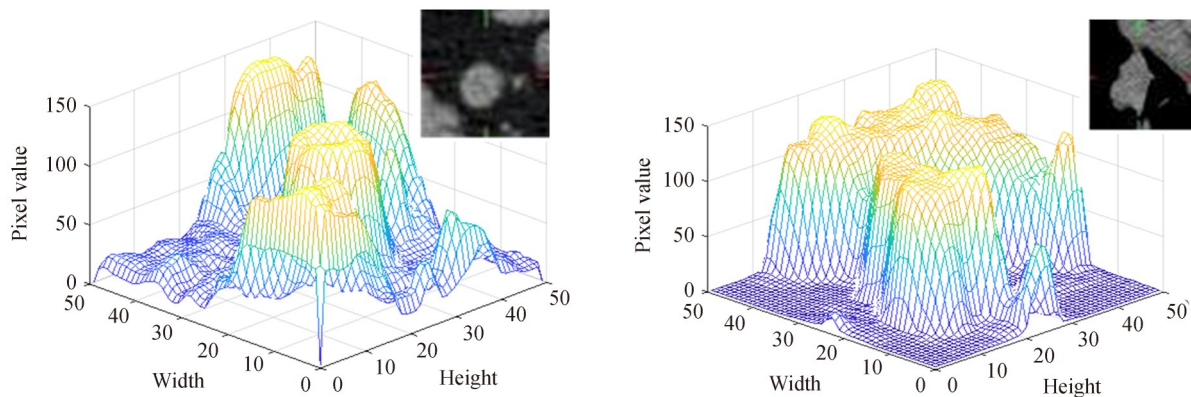


图6 CT图像三维灰度图
Fig.6 CT image three-dimensional gray value

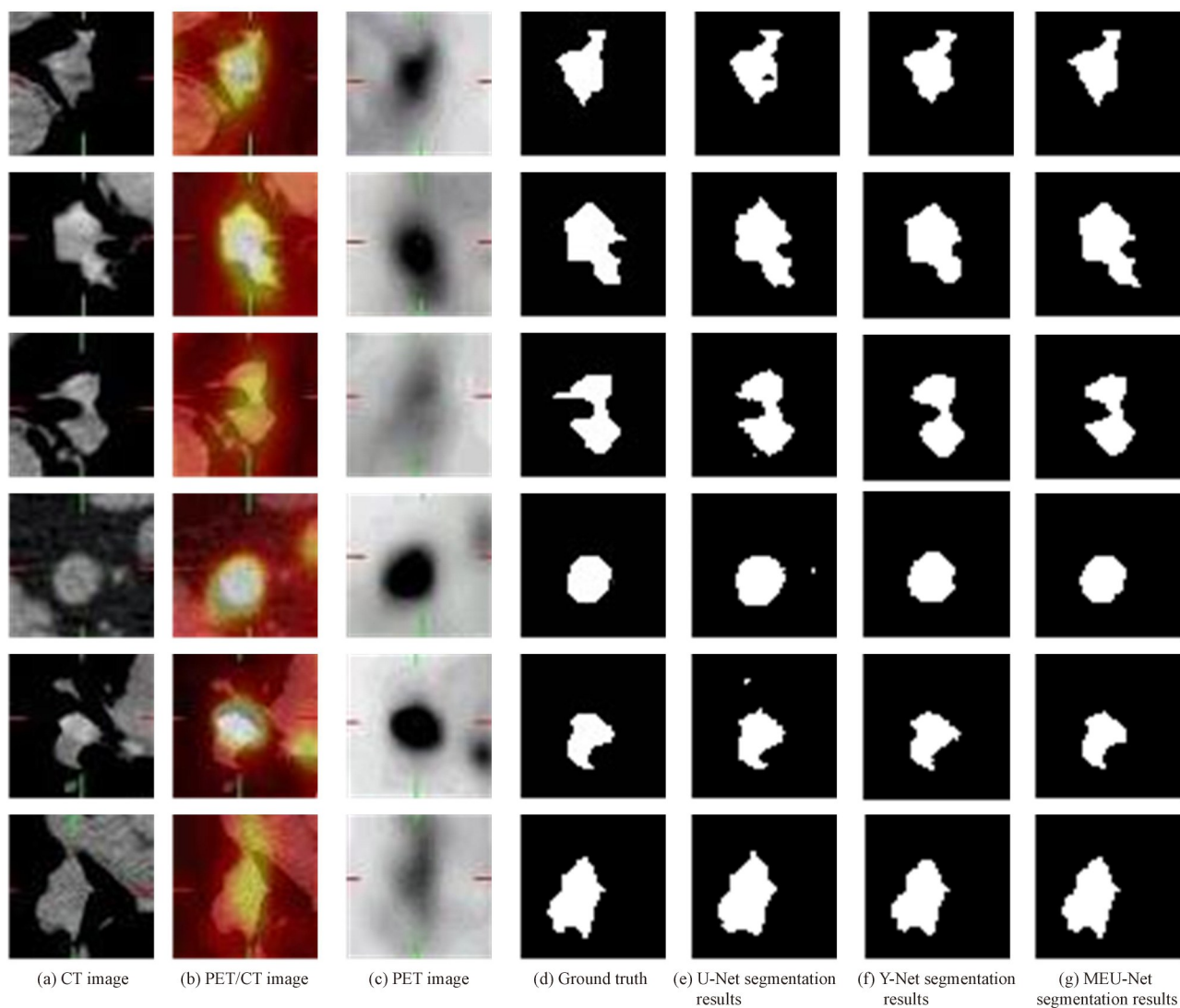


图7 不同编码器网络分割结果
Fig.7 Network segmentation results of different encoders

表 1 多编码器分割结果
Table 1 Segmentation results of multi-encoders

Architecture	DSC/%	Recall/%	VOE/%	RVD/%
U-Net ^[8]	95.16	94.99	92.53	92.74
Y-Net ^[24]	95.13	95.0	92.52	92.7
MEU-Net	95.20	95.13	92.59	92.76

从表 1 中可以看出,两编码器 Y-Net 网络的 DSC 和 Recall 指标不如 U-Net 网络,分别降低了 0.03% 和 0.01%,而 VOE 和 RVD 指标比 U-Net 好,分别增加了 0.01% 和 0.04%,两编码器 Y-Net 网络输入 PET 和 CT 图像提取特征,由于 PET 图像不同于 CT 图像,其提供病灶的代谢信息,帮助定位病灶,但是缺乏清晰的病灶轮廓信息,所以仅部分指标高于 U-Net。三编码器 MEU-Net 的 DSC 和 Recall 较两编码器 Y-Net 分别提升了 0.07% 和 0.13%,VOE 和 RVD 分别增加了 0.07% 和 0.06%,总的来说,三编码器 MEU-Net 网络的分割指标均高于两编码器 Y-Net 和 U-Net,DSC、Recall、VOE 和 RVD 分别为 95.20%、95.13%、92.59% 和 92.76%,由此可以看出模态互补的医学图像对病灶分割性能有一定提升。图 7 为不同编码器网络的分割结果,从图中可以看出,肺部病灶形状复杂且与正常组织粘连,具有一定的分割难度,U-Net 分割部分病灶出现了欠分割情况,如图 7(e) 第一行。两编码器 Y-Net^[24] 对于病灶形状复杂的分割效果不如其它两个网络,如图 7(f) 的第二行和第三行。三编码器 MEU-Net 网络对于轮廓复杂与正常组织粘连的病灶分割效果优于其它两个网络,如图 7(g) 的第二、三行。图 8 使用雷达图比较不同分割架构的性能,可以看出三编码器 MEU-Net 的 DSC、Recall、VOE 和 RVD 均比其它网络架构高,说明其效果优于其它两个网络。

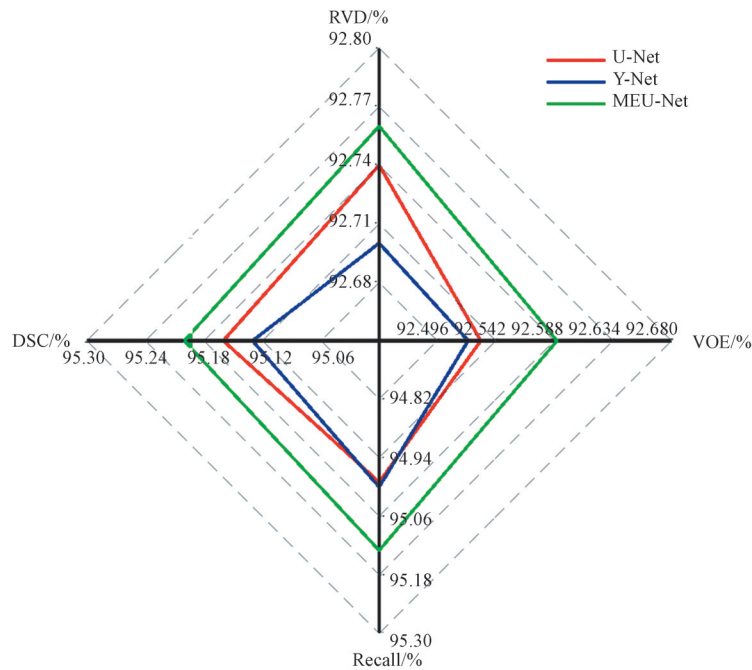


图 8 不同编码器分割结果对比

Fig.8 Comparison of segmentation results of different encoders

2.4.2 混合注意力机制

本组实验基于三编码器 U-Net 评估混合注意力机制的性能,共五次实验,实验模型均使用 CT、PET/CT 和 PET 图像各 909 张作训练集,CT、PET/CT 和 PET 图像各 117 张作测试集。实验一,模型跳跃连接中没有使用任何注意力机制,即为 MEU-Net;实验二,MEU-Net 模型跳跃连接只引入空间注意力机制,模型为 Multi Encoder Spatial Attention U-Net (MESAU-Net);实验三,模型跳跃连接只引入通道注意力机制,模型为 Multi Encoder Channel Attention U-Net (MECAU-Net);实验四,模型跳跃连接引入本文空间注意力机制和文献[25]中的轻量级通道注意力机制,模型为 MEAU-Net[#],该通道注意力模块为无需降维的局部跨通道互动策略保留 Squeeze and Excitation Networks (SENet)^[26] 的全局平均池化,使用自适应核的一维卷

积代替SENet的两个全连接层;实验五,模型跳跃连接引入本文的混合注意力机制,模型为MEAU-Net。图9为CT图像的三维灰度图,表2为不同注意力机制的评价指标,图10为不同注意力机制模型分割结果。

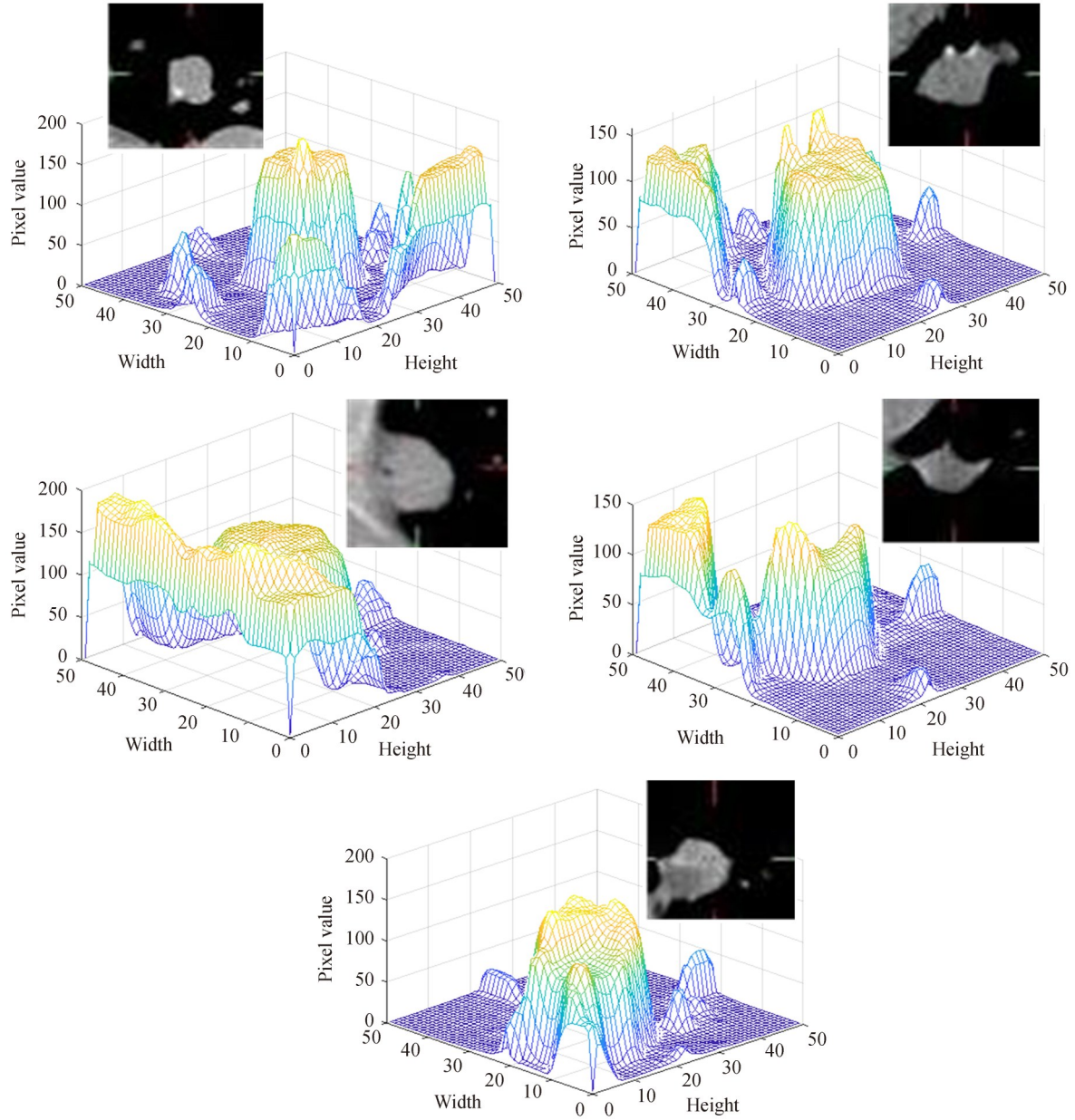


图9 CT图像三维灰度图

Fig.9 CT image three-dimensional gray value

从表2可以看出模型MECAU-Net和MESAU-Net的各项指标均高于三编码器U-Net模型MEU-Net,说明网络跳跃连接引入空间注意力机制和通道注意力机制能提高网络分割性能。其中引入通道注意

表2 不同注意力机制分割结果

Table 2 Segment results of different attention mechanisms

Architecture	DSC/%	Recall/%	VOE/%	RVD/%
MEU-Net	95.20	95.13	92.59	92.76
MESAU-Net	95.43	96.26	92.76	92.7
MECAU-Net	95.68	96.33	92.65	92.8
MEAU-Net [#]	96.0	96.5	92.7	92.68
MEAU-Net	96.4	97.27	93.0	93.06

力机制的MECAU-Net大多数指标优于引入空间注意力机制MESAU-Net的指标,DSC、Recall和VOE分别提升了0.25%、0.07%、0.1%。MEAU-Net[#]的部分指标优于引入单注意力机制的网络,如DSC、Recall。混合注意力机制MEAU-Net的DSC、Recall、VOE和RVD分别为96.4%、97.27%、93.0%和93.06%。从图10可以看出,MEAU-Net[#]网络的分割结果整体优于前三种方法,但部分分割效果不佳,如图10(h)第一行出现了欠分割情况,第二行病灶边缘分割模糊。从图10第四、五行可以看出,混合注意力机制网络MEAU-Net对于正常组织粘连、形状复杂的病灶分割效果优于其他方法。从图11可以看出,MEAU-Net的指标优

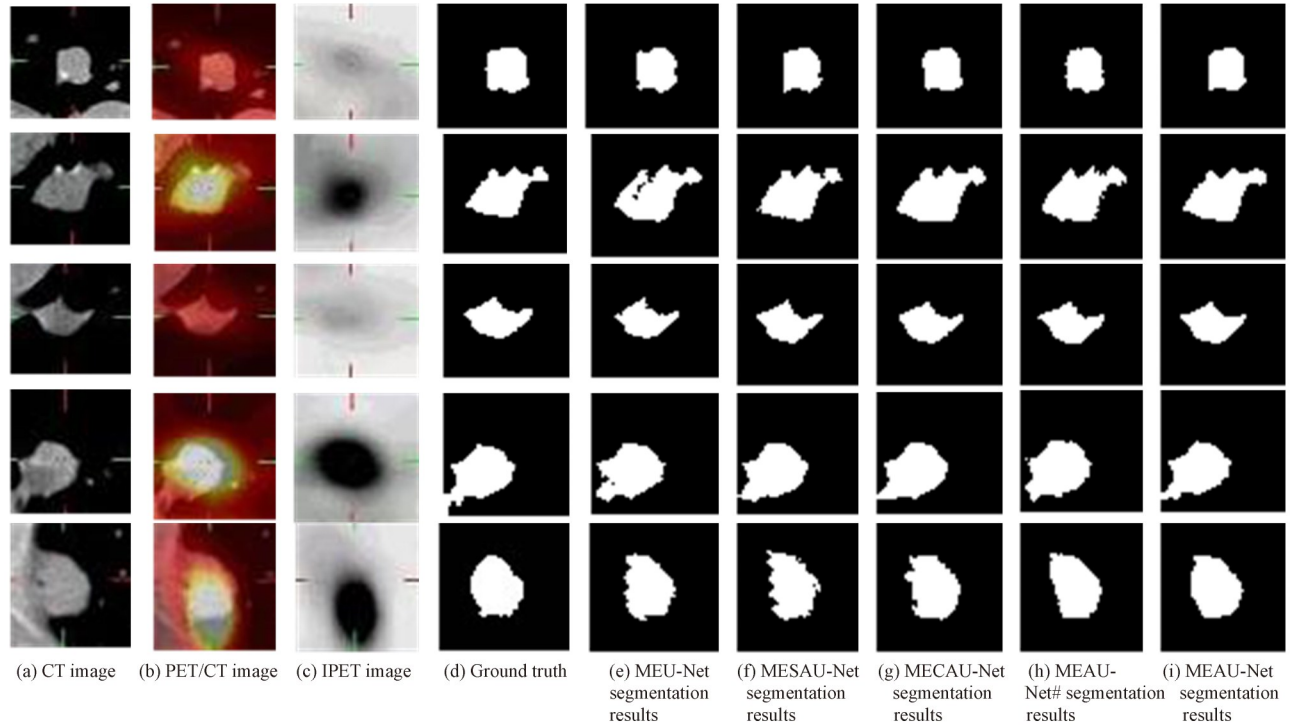


图10 不同注意力机制网络分割结果

Fig.10 Network segmentation results of different attention mechanisms

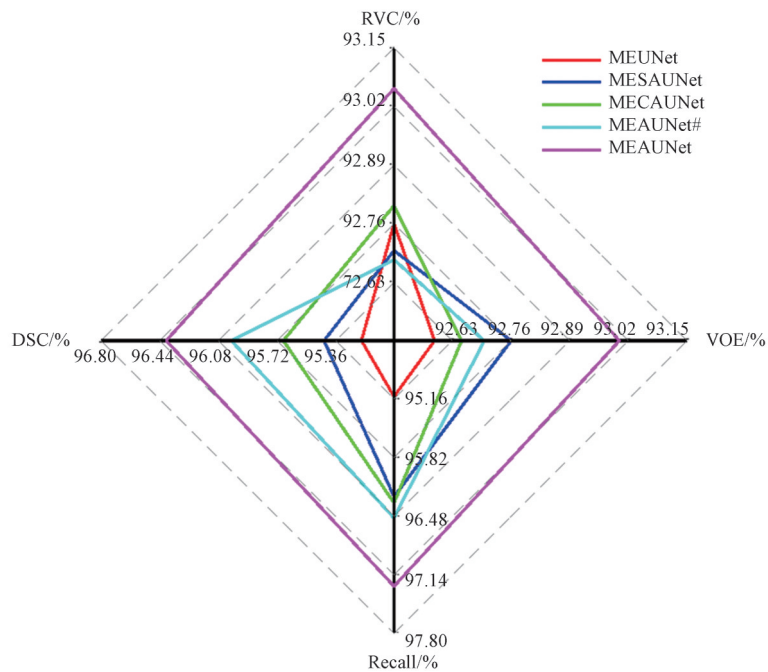


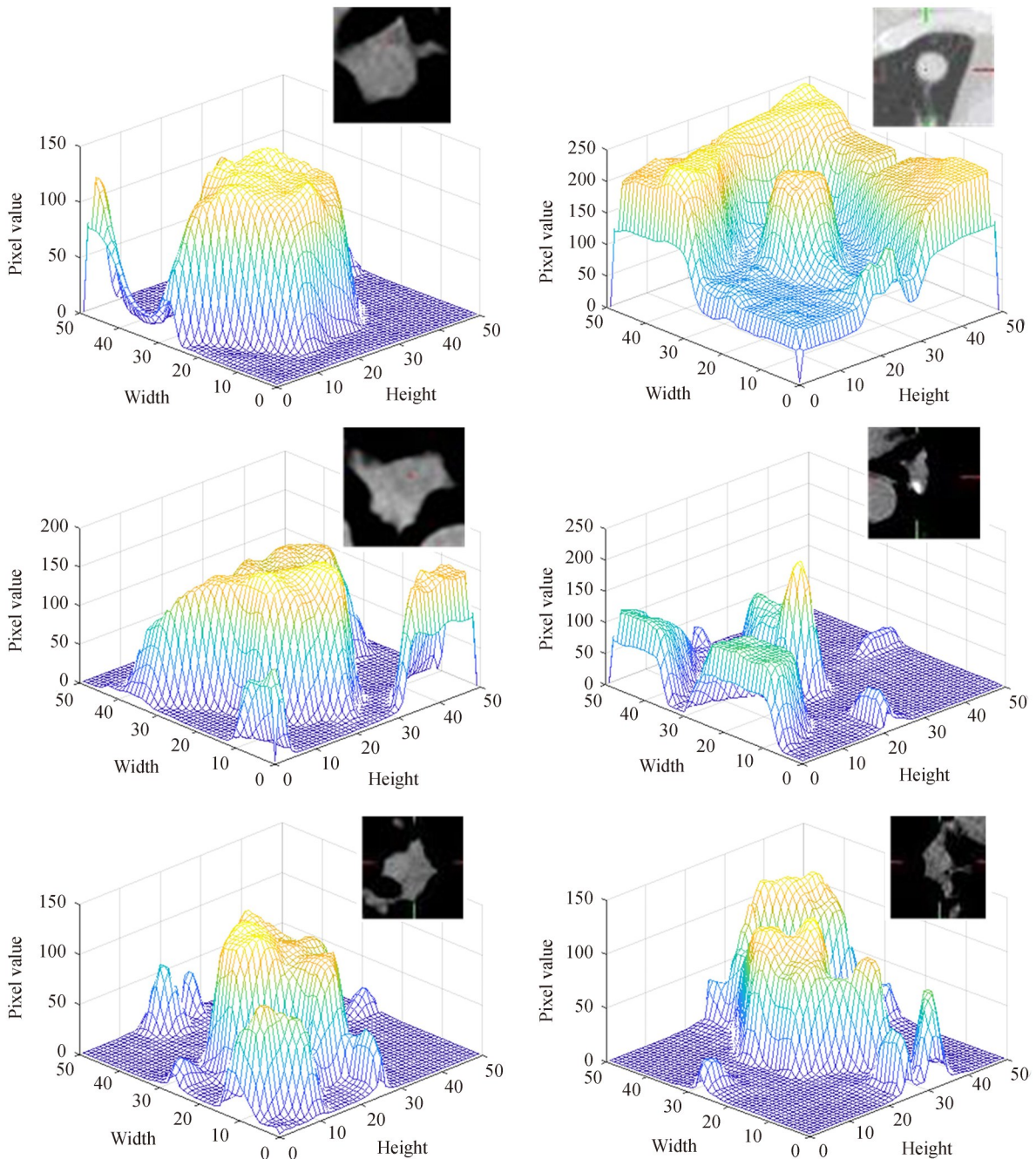
图11 不同注意力机制的分割结果对比

Fig.11 Comparison of segmentation results of different attention mechanisms

于其他注意力机制方法,混合注意力机制对网络分割性能有一定提升效果,说明在跳跃连接过程中引入混合注意力机制能有效提升分割效果。

2.4.3 与先进分割算法对比

本文方法与编解码器网络、多模态医学图像分割方法和注意力机制相关方法进行对比。其中SegNet^[27]的编码网络为VGG-16的卷积层,网络跳跃连接部分使用池化位置索引存储每个池化窗口中最大特征值的位置代替U-Net的拼接操作,提供分割效率。Wnet^[28]是级连的两个U-Net网络,用于PET/CT图像骨髓瘤的全身骨病变分割,第一个U-Net网络层数为5层,输入CT图像,第二个U-Net网络层数为3层,该子网络使用第一个U-Net分割结果和PET图像作为第二个网络的输入。Attention U-Net^[29]在U-Net的基础上,在网络的跳跃连接中嵌入自注意力门,将网络上采样的特征和编码路径传输尺寸相同的特征通过自注意力门自动学习分割目标的尺寸和外形,用于胰腺CT图像病灶分割。CT图像的三维灰度图如图12所示,MEAU-Net与其他先进算法的对比结果如表3所示,分割结果如图13所示。



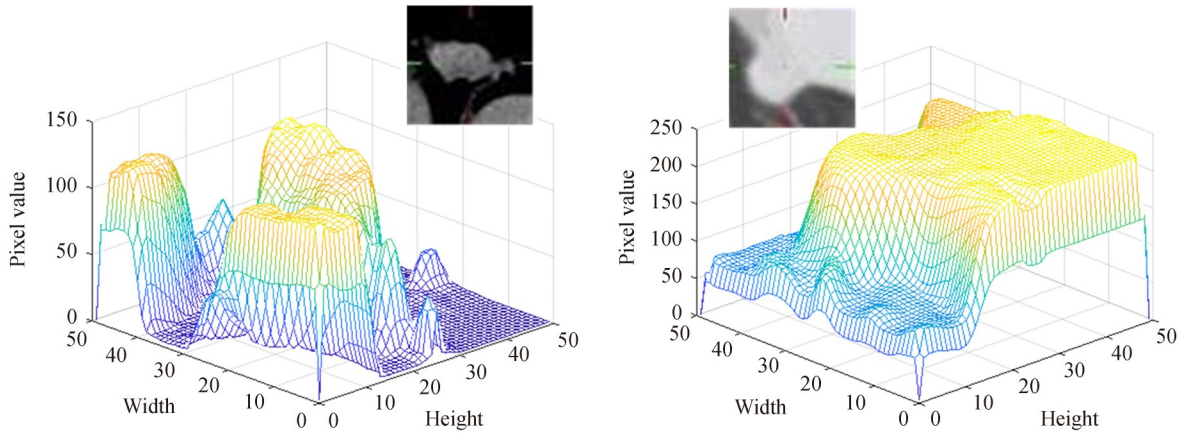


图12 CT图像的三维灰度图
Fig.12 CT image three-dimensional gray value

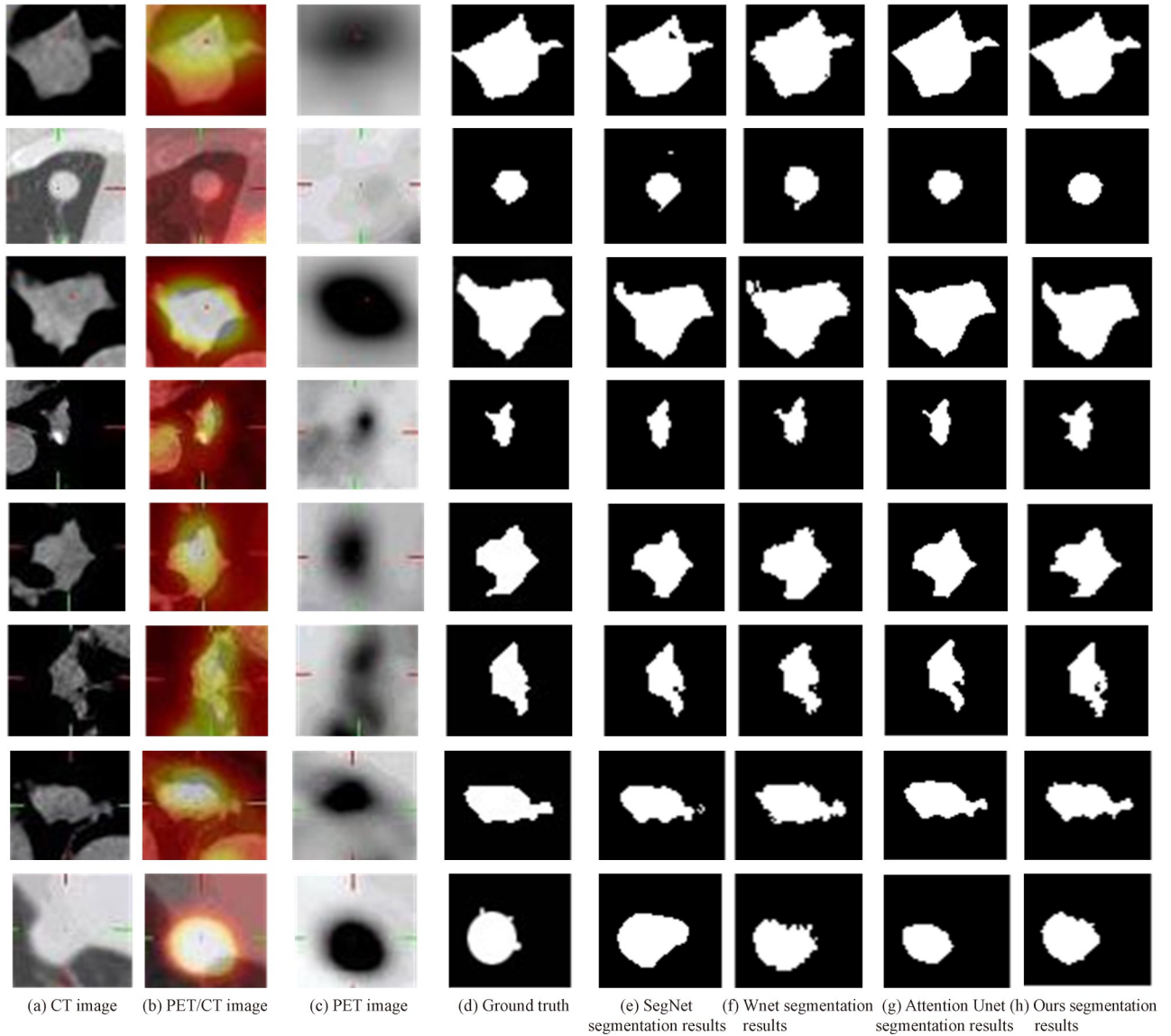


图13 不同算法分割结果
Fig.13 Segmentation results of different methods

表3 MEAU-Net与其它网络的分割结果
Table 3 Segmentation results of MEAU-Net and other networks

Architecture	DSC/%	Recall/%	VOE/%	RVD/%
SegNet ^[27]	94.82	95.11	91.81	92.04
Wnet ^[28]	94.73	95.98	92.08	92.17
Attention Unet ^[29]	95.69	96.17	92.64	92.73
Ours	96.4	97.27	93.0	93.06

从表3可得, SegNet的DSC、Recall、VOE和RVD分别为94.82%、95.11%、91.81%和92.04%, 该网络仅使用CT单模态图像进行病灶分割, 对于与正常组织粘连的病灶, 分割效果不佳。从图13可以看出, SegNet分割结果出现了欠分割情况如图13(e)第一行, 对于与正常组织粘连的病灶出现了过分割情况, 如图13(e)最后一行。Wnet的DSC、Recall、VOE、RVD分别94.73%、95.98%、92.08%和92.17%, 大部分指标高于SegNet, 如Recall、VOE和RVD, 该方法将第一个网络得到的CT特征图输入PET图像, 充分利用CT图像的细节信息, PET图像提供了病灶的代谢信息, 有助于定位病灶, 但由于第二个网络层数过浅未能充分学习特征, 并且PET图像中病灶模糊的边缘信息导致分割效果低于引入注意力机制的网络。图13(f)第二行该网络出现了欠分割的情况, 对于第三、七行形状复杂的病灶, 分割效果不佳。Attention Unet的DSC、Recall、VOE、RVD分别为95.69%、96.17%、92.64%和92.73%, 相较于Wnet分别提高了0.96%、0.19%、0.56%和0.56%。由于Attention Unet在网络跳跃连接部分使用注意力门自动关注病灶区域, 因此分割效果优于SegNet、Wnet, 但是对于形状复杂的病灶如图13(a)第6行和与正常组织粘连的病灶如图13(a)最后一行, 没有PET图像提供的代谢信息, Attention Unet分割效果不如MEAU-Net。MEAU-Net的DSC、Recall、VOE、RVD分别为96.4%、97.27%、93.0%和93.06%, 图13(h)中可以看出, MEAU-Net对于不同尺度的病灶均能有效分割, 对于与正常组织粘连的病灶, MEAU-Net通过多模态图像之间的特征互补, 对病灶精确定位, 分割结果比其他方法好。图14为不同分割方法的指标对比结果, 从图中可以看出, MEAU-Net的各项指标优于其他网络。SegNet和Attention Unet方法都未使用PET、CT和PET/CT多模态医学图像特征, 仅利用单模态医学图像, 忽略了多模态医学图像之间的优势互补。由此证明MEAU-Net使用多编码器充分提取多模态医学图像特征是可行的, 混合注意力机制选取重要特征, 聚焦在特征图中病灶位置, 能够

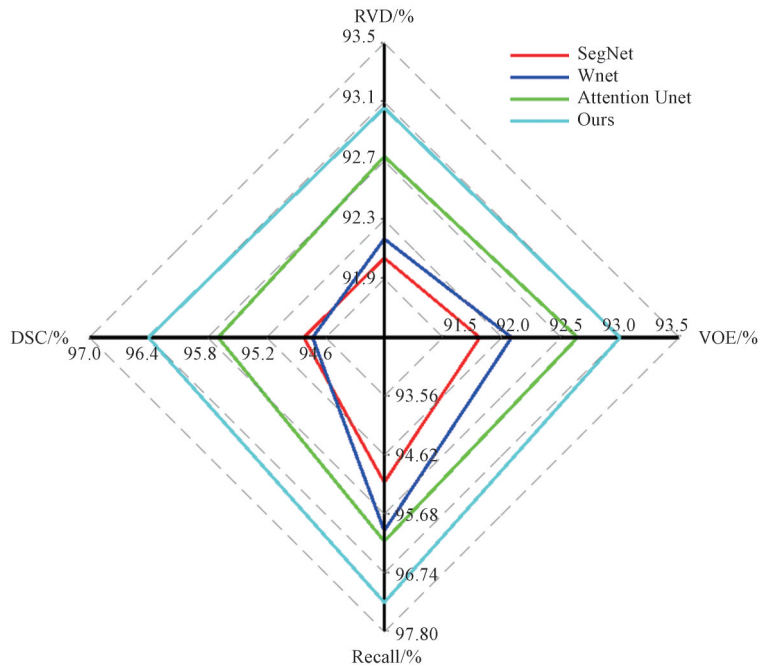


图14 不同方法分割结果对比
Fig.14 Comparison of segmentation results of different methods

提高网络分割性能,为医学分割提供更准确的判断依据,分割结果更接近标签值。

3 结论

本文充分利用不同医学成像设备成像机理的特点,提出了多编码混合注意力的分割网络MEAU-Net。使用多编码器提取多模态医学图像肺部病灶特征,通过在网络跳跃连接引入混合注意力机制,使网络自动关注在图像病灶区域和重要特征通道。对于解码路径包含的不同尺度高级语义特征,构建了多尺度特征聚合块,实现了不同尺度的特征聚合。在临床数据集上对该算法进行验证,实验结果表明MEAU-Net的DSC、Recall、VOE和RVD的平均值分别为96.4%、97.27%、93.0%和93.06%。相较于三种对比算法,MEAU-Net对于形状复杂、病灶与正常组织相粘连的情况均能有效分割。接下来将进一步扩充数据集,并将方法推广到3D肺部肿瘤的分割和其它疾病的精准分割。

参考文献

- [1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics[J]. CA: A Cancer Journal for Clinicians, 2018, 68: 7-30.
- [2] GOPI K, SELVAKUMAR J, AMIR H G, et al. Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier[J]. Expert Systems with Applications, 2019, 134: 112-119.
- [3] KAUSHAL C, BHAT S, KOUNDAL D, et al. Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images[J]. IRBM, 2019, 40(4): 211-227.
- [4] JAMSHID S N, ABDOLLAH K, BEHROOZ N. Lung tumor segmentation using improved region growing algorithm[J]. Nuclear Engineering and Technology, 2020, 52(10): 2313-2319.
- [5] YANG Y, CHONG F, WANG R. Automatic segmentation model combining U-Net and level set method for medical images[J]. Expert Systems with Applications, 2020, 153: 113419.
- [6] SHOUVIK C, KALYANI M. A morphology-based radiological image segmentation approach for efficient screening of COVID-19[J]. Biomedical Signal Processing and Control, 2021, 69: 102800.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]. The International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich Germany, 2015: 234-241.
- [8] LIU L, CHENG J, QUAN Q, et al. A survey on U-shaped networks in medical image segmentations [J]. Neurocomputing, 2020, 409: 244-258.
- [9] LI Daxiang, ZHANG zhen. Improved U-Net segmentation algorithm for the retinal blood vessel images[J]. Acta Optica Sinica, 2020, 40(10): 1010001.
李大湘,张振.基于改进U-Net视网膜血管图像分割算法[J].光学学报,2020,40(10):1010001.
- [10] HUANG Hong, LV Rongfei, TAO Junli, et al. Segmentation of lung nodules in CT images using improved U-Net++ [J]. Acta Photonica Sinica, 2021, 50(2): 0210001.
黄鸿,吕容飞,陶俊利,等.基于改进U-Net++的CT影像肺结节分割算法[J].光子学报,2021,50(2):0210001.
- [11] ZHANG J, LI C, KOSOV S, et al. LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation[J]. Pattern Recognition, 2021, 115: 107885.
- [12] GUO F, CAPALDI P D, MCCORMACK D G, et al. Ultra-short echo-time magnetic resonance imaging lung segmentation with under-annotations and domain shift[J]. Medical Image Analysis, 2021, 72: 102107.
- [13] ZHOU Tao, DONG Yali, HUO Bingqiang, et al. U-Net and its applications in medical image segmentation: a review [J]. Journal of Image and Graphics, 2021, 26(9): 2058-2077.
周涛,董雅丽,霍兵强,等.U-Net网络医学图像分割应用综述[J].中国图象图形学报,2021,26(9):2058-2077.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]. European Conference on Computer Vision-ECCV, Springer, Cham, 2018: 3-19.
- [15] GU R, WANG G, SONG T, et al. CA-Net: comprehensive attention convolutional neural networks for explainable medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(2): 699-711.
- [16] REYHANEH M F, NAJME K Z, REZA V, et al. 2-[18F]FDG PET/CT radiomics in lung cancer: an overview of the technical aspect and its emerging role in management of the disease[J]. Methods, 2021, 188: 84-97.
- [17] GUO Y, ZHU H, YAO Z, et al. The diagnostic and predictive efficacy of 18F-FDG PET/CT metabolic parameters for EGFR mutation status in non-small-cell lung cancer: a meta-analysis [J]. European Journal of Radiology, 2021, 141: 109792.
- [18] JULIAN M M R, NIKOLAJ F, STEPHANIE B, et al. FDG-PET/CT for pretherapeutic lymph node staging in non-small cell lung cancer: a tailored approach to the ESTS/ESMO guideline workflow [J]. Lung Cancer, 2021, 157: 66-74.
- [19] HISASHI O, AKIRA S, HIROTSUGU N, et al. Correlation between preoperative 18F-FDG PET/CT findings and

- postoperative short-term prognosis in lung cancer patients with idiopathic interstitial pneumonia after lung resection [J]. *Respiratory Investigation*, 2021, 59(1):106-113.
- [20] SARA S, FRANCO V, RUSSELL K H, et al. Imaging in therapy response assessment and surveillance of lung cancer: evidenced-based review with focus on the utility of 18F-FDG PET/CT[J]. *Clinical Lung Cancer*, 2020, 21(6):485-497.
- [21] ZHOU Tao, LU Huiling, WEI Xingyu, et al. Self-adaption fusion algorithm for lung cancer PET/CT based on Piella frame and DT-CWT[J]. *Journal of University of Science and Technology of China*, 2017, 47(1):10-17.
周涛, 陆惠玲, 魏兴瑜, 等. 基于 Piella 框架和 DT-CWT 的肺癌 PET/CT 自适应融合算法[J]. *中国科学技术大学学报*, 2017, 47(1):10-17.
- [22] TAKAHIRO Y, TSUKASA M. Detection and localization of manhole and joint covers in radar images by support vector machine and Hough transform[J]. *Automation in Construction*, 2021, 126:103651.
- [23] YING Z, GE L, REN Y, et al. A new image contrast enhancement algorithm using exposure fusion framework [J]. *Computer Analysis of Images and Patterns*, 2017, 10425:36-46.
- [24] LAN H, JIANG D, YANG C, et al. Y-Net: hybrid deep learning image reconstruction for photoacoustic tomography in vivo[J]. *Photoacoustics*, 2020, 20:100197.
- [25] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]. *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA. 2020:11531-11539.
- [26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA. 2018:7132-7141.
- [27] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12):2481-2495.
- [28] XU L, GILES T, JANA L, et al. Automated whole-body bone lesion detection for multiple myeloma on 68Ga-pentixafor PET/CT imaging using deep learning methods[J]. *Contrast Media & Molecular Imaging*, 2018, 2018:1-11.
- [29] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[OL]. <https://arxiv.org/pdf/1804.03999v3.pdf>. 2021.8.26.

Cross-modality Multi-encoder Hybrid Attention U-Net for Lung Tumors Images Segmentation

ZHOU Tao^{1,2}, DONG Yali¹, LIU Shan¹, LU Huiling³, MA Zongjun⁴, HOU Senbao¹, QIU Shi⁵

(1 School of Computer Science and Technology, North Minzu University, Yinchuan 750021, China)

(2 The Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China)

(3 School of Science, Ningxia Medical University, Yinchuan 750004, China)

(4 Department of Orthopedics, Ningxia Medical University General Hospital, Yinchuan 750004, China)

(5 Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China)

Abstract: The lung lesions segmentation in medical imaging is an important task. However, there are still some challenges. The lesions delineation relies on manual segmentation by experienced clinicians, which is time-consuming and labor-intensive due to the complex anatomical structure of the human body; Lung tumor images have the characteristics of low contrast, different size and shape of the lesions, and variable location of the lesions, and are characterized by unbalanced data distribution. U-Net can segment lesions under a small number of datasets and has been widely used in medical image segmentation of lesions and organs. However, U-Net has the following three problems. First, U-Net uses uniform parameters for each feature map. For lesions of different sizes and complex shapes, the network may have poor spatial perception, which leads to the decline of segmentation performance. Second, U-Net channel dimension doubles with the number of down-sampling, and the feature map of the encoder layer is concatenated to the decoding layer through skip connection. However, in the segmentation task, the importance of different channels to the segmentation task is different. Third, most of the current multi-encoder segmentation networks extract the features of the single-modal target slice and their continuous slices to improve the network segmentation performance, but ignore the ability of different modal medical images to express the characteristics of the lesion. To solve the above problems, this paper proposes the MEAU-Net network to extract complementary features of multi-modals images. First, for the unbalanced data distribution, the

Hough transform is used to detect the line of the lung Computed Tomography (CT) image marked by the doctor to obtain the region of interest, and cropped image size from 356 pixel \times 356 pixel to 50 pixel \times 50 pixel. Then, for the low contrast of medical image, use exposure fusion image contrast enhancement method improves the contrast between lesion and the background of lung CT image. To extract the features of lesions in multi-modal medical images, this paper proposes a multi-encoder hybrid attention mechanism network MEAU-Net. Positron Emission Tomography (PET) images provide metabolic information of lesions, CT images provide anatomical information of lesions, and Positron Emission Tomography/Computed Tomography (PET/CT) images combine their advantages and utilize their complementarity and redundancy. MEAU-Net encoder path includes three branches of PET/CT, PET and CT, which are used to extract corresponding modal image features. In the skip connection of the network, hybrid attention mechanism is used, including spatial attention mechanism and channel attention mechanism. The features of PET/CT and CT are used in the spatial attention mechanism to emphasize key areas in the feature map and suppress irrelevant background. The channel attention mechanism extracts the weight value of each channel for the three branches of PET/CT, CT and PET, and then selects the maximum weight value after the three branches sigmoid to multiply the corresponding channel, and assign a higher value to the important channel. The weighting coefficient realizes the selection of important channels. The network inputs the feature map through the hybrid attention mechanism into the corresponding decoder layer, so that the network focuses on the lesion part in medical image, suppresses useless background information, and achieves accurate segmentation of the image lesion. Finally, for the semantic features of different scales of the decoding path, this paper uses a multi-scale feature aggregation block to perform feature mapping on the features of the decoding path, and refine the segmentation of the lesion. We compared our model with 4 classical segmentation model on our dataset, including SegNet, Attention Unet and Wnet. The experiment results show that our model uses multi-modal medical image features to effectively segment lung lesions with complex shapes, and outperforms all other methods in our dataset. The average DSC, Recall, VOE and RVD of MEAU-Net segmentation results are 96.4%, 97.27%, 7.0% and 6.94%, respectively.

Key words: Deep learning; Medical image segmentation; Multimodal medical image; U-Net; Lung tumor
OCIS Codes: 100.2960; 100.4996; 170.3880