

分类号 TP391.4

密级

UDC

编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕 士 学 位 论 文

用于全身 PET/CT 肿瘤分割的位置监督的
多分支神经网络

研 究 生 姓 名: 王 堃

指导教师姓名、职称: 何江萍 副教授

学 科、专 业 名 称: 管理科学与工程

研 究 方 向: 信息管理与信息系统

提 交 日 期: 2021 年 5 月 24 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 王堃 签字日期： 2021.5.24

导师签名： 何江萍 签字日期： 2021.5.24

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 王堃 签字日期： 2021.5.24

导师签名： 何江萍 签字日期： 2021.5.24

Metabolic Tumor Segmentation from Whole-body PET/CT Images Using a Position Supervised Multi-branch Neural Network

Candidate: Wang Kun

Supervisor: He Jiangping

摘要

肺癌是世界第二大常见癌症，且在所有癌症中致死率最高。全身肿瘤体积（whole-body metabolic tumor volume, MTV_{wb} ）是用于预测肺癌患者生存时间的独立变量，目前该变量的计算依赖医生手动勾画的肿瘤标签。勾画肿瘤标签是一个费时费力、重复性差的工作，而且医生在勾画时，由于肿瘤边界的模糊性，会产生主观差异。使用计算机自动分割肿瘤能辅助医生对癌症的诊断。现有 PET/CT 肺癌肿瘤分割方法在胸部图像表现较好，但没有考虑肿瘤向全身转移的情况。由于全身图像远比胸部图像复杂，直接将这些方法套用在全身图像上并不能取得很好的效果，因此需要针对全身 PET/CT 的肿瘤分割方法。

本文提出了一种位置监督的多分支神经网络（position supervised multi-branch network, PSMB-Net）用于从全身 PET/CT 中分割肿瘤。方法从三个角度进行研究：第一，CT 图像包含了人体固有的结构规律，CT 切片稳定的相对位置体现了这种规律，对该相对位置的学习可以辅助 PET/CT 肿瘤分割；第二，全身 PET 图像中存在与肿瘤高度相似的正常代谢活跃组织，胸外 CT 中肿瘤与周围正常组织的边界极难划分，使得全身图像的规律远比胸部图像复杂。通过观察发现全身 PET/CT 可以以肺部下界为标准，分为上下两个半身，使得每个半身图像内部的规律比全身图像更为简单。基于此，本文定义了两个基于位置的解码器，分别学习人体上下两个半身的特征，用两个门信号控制它们的优化过程，门信号将切片的相对位置转换为切片在输出中的权重；第三，全身 PET/CT 图像大肿瘤与小肿瘤的像素个数差距悬殊，使得在标准的交叉熵分割损失的优化下，模型对小肿瘤的预测不够全面。在标准交叉熵中加入一个肿瘤级的平衡系数可以缓解这个问题。

实验使用 480 位肺癌患者的全身 PET/CT 测试该方法的性能。在 5 折交叉测试中，分别计算图像分割指标和生存分析指标。本文提出的 PSMB-Net 在 Dice、Sensitivity、Precision 上分别达到了 0.580、0.616 和 0.688，均超过了所比较的方法。在生存分析中，PSMB-Net 预测的 MTV_{wb} 计算出的 C-index 比来自医生勾画的 MTV_{wb} 在单变量下的 Cox 和 RSF 模型上仅低 0.023 和 0.008。这些结果说明本文所提方法在全身 PET/CT 肿瘤分割任务上有一定价值。

关键词：肿瘤分割 全身 PET/CT 相对位置 生存分析

Abstract

Lung cancer is on the second rank of most common form of internal malignancies, and it is the leading cause of cancer related death in 2020. Whole-body metabolic tumor volume (MTV_{wb}) is an independent prognostic factor for overall survival in lung cancer patients. Current calculation of this factor is depended on the tumor label provided by radiologists, which is labor consuming with poor reproducibility, and objective difference is introduced during the process of label delineation due to the blurry border. Automatic segmentation of lung cancer tumors to calculate MTV_{wb} can assist radiologists in their assessment of lung cancer. Published methods appeared promising in segmentation accuracy in thoracic PET/CT, their methods do not quantify distant metastatic disease outside of the thoracic region. A simple application of these methods on whole-body PET/CT is not sufficient to generate acceptable results due to the greatly increased complexity, thus it is essential to search of a method for whole-body PET/CT segmentation.

We present a position supervised multi-branch neural network (PSMB-Net) to segment tumor from whole-body PET/CT images. We make our research from three points of view. First, CT images contain inherent structural pattern, and the relative position of CT slices is a form of this pattern. The learning of the relative position can assist tumor segmentation of PET/CT images. Second, normal organs and tissues can be metabolically active, and therefore appear similar to tumors on whole-body PET images, and the border of extra-thoracic tumors are difficult to identify on CT images, making the complexity of whole-body PET/CT is much higher than thoracic image. To accommodate the remarkable visual differences of the upper and lower semi-body, which is determined by the lower bound of lung, our design includes two decoders to learn features

from the upper and lower semi-body separately. Two position-based gate signals transfer each position value to the weight of the corresponding slice in the output. Third, the voxel number of greater tumor versus smaller tumor are in greatly imbalanced, making the model biased to ignore smaller tumors in test stage, under the optimization of standard cross-entropy segmentation loss. We add a balance factor to cross-entropy loss function to alleviate this imbalance.

We test the performance of our method on a PET/CT image dataset including 480 lung cancer patients with 5-fold cross-validation using the metrics for image segmentation and survival analysis. Our PSMB-Net archived 0.580, 0.616, and 0.688, on Dice, Sensitivity and Precision, respectively, which archives higher performance than some state-of-art methods. The concordance indices (C-indices) for the univariate MTV_{wb} segmented by PSMB-Net are only 0.023 and 0.008 lower than that calculated from tumor label, in Cox and RSF model, respectively. These results demonstrate that our method is of value for whole-body tumor segmentation of PET/CT images.

Keywords: Tumor segmentation; Whole-body PET/CT; Relative position ; Survival analysis

目 录

1 引言	1
1.1 选题依据	1
1.2 PET/CT 肿瘤分割方法研究现状	2
1.2.1 在 PET 图像上分割的方法.....	2
1.2.2 在 PET/CT 图像上的分割方法.....	7
1.3 本文研究内容.....	9
1.4 本文组织结构.....	9
2 卷积神经网络	11
2.1 概念.....	11
2.2 基本结构.....	13
2.2.1 卷积层.....	13
2.2.2 池化.....	17
2.2.3 激活函数.....	18
2.2.4 损失函数.....	19
2.3 训练过程.....	21
2.3.1 参数初始化.....	21
2.3.2 学习率.....	23
2.3.3 数据预处理.....	23
2.3.4 反向传播.....	24
2.4 本章总结.....	25
3 基于 PSMB-Net 的 PET/CT 肿瘤分割	26
3.1 模型结构.....	26
3.1.1 CT 切片的位置监督	26
3.1.2 CT 特征融合	27
3.1.3 门信号生成.....	28
3.1.4 模态特征融合与肿瘤分割.....	29

3.2 相对位置计算.....	30
3.3 肿瘤级的平衡系数.....	31
3.4 损失函数.....	32
3.5 本章总结.....	32
4 实验.....	34
4.1 数据.....	34
4.2 实现细节.....	35
4.3 其它方法的实现.....	36
4.4 结果.....	37
4.4.1 指标.....	37
4.4.2 比较和分析.....	38
4.5 讨论.....	42
4.6 本章总结.....	44
5 展望与总结	46
5.1 展望.....	46
5.2 总结.....	46
参考文献	48
致谢.....	57
硕士期间科研项目与成果.....	58

1 引言

本章介绍了本文的研究背景和全身肿瘤自动分割在临床辅助决策中的重要意义，在此基础上介绍了全身肿瘤分割相比于局部肿瘤分割的难点。其次，本章介绍了 PET/CT 肿瘤分割方法的现有研究，针对各种方法总结了优点和不足。最后，对本文的组织结构做了简要介绍。

1.1 选题依据

PET/CT 是一种广泛使用的多模态医学影像，它用于癌症患者的治疗计划和预后评估。氟代脱氧葡萄糖正电子发射断层成像（fluorodeoxyglucose-positron emission tomography, FDG-PET）通过使用放射性追踪剂 18F-氟代脱氧葡萄糖（18F-fluorodeoxyglucose, 18F-FDG）来检测组织活性，癌症组织会在其图像中呈现高于正常组织的吸收量。但是，由于 FDG-PET 图像的分辨率低，细节不能很好地反映出来。计算机断层成像（computed tomography, CT）是一种基于组织密度的成像模态，其图像的分辨率高。PET/CT 的结合使得人们在观测组织活跃性的同时确定其对应的解剖结构，在人体的各个部位检测癌症。

根据 2020 年全球癌症统计，肺癌是世界第二大常见的癌症，占有癌症病例的 11.4%，并且在所有因癌症死亡的病例中以 18.0% 的比例位列第一^[1]。因此本文把肺癌患者作为研究对象。

全身肿瘤体积（whole-body metabolic tumor volume, MTV_{wb} ）指患者体内肿瘤的总容积。研究证明 MTV_{wb} 是重要的预测肺癌患者的生存时间的变量^{[2][3]}。此外，肿瘤的临床分期、疗效评价等都依赖 MTV_{wb} 。但在目前的临床实践中，PET/CT 图像中的肿瘤是用手动勾画的方式分割的，这是一个费时费力且可重复性差的工作。因此，肿瘤的自动分割方法有辅助放射科医生更高效地诊断癌症的价值。

Park 等人发现 79% 的 4 期非小细胞肺癌患者有胸外转移的肿瘤^[4]，但是现有的肺癌 PET/CT 肿瘤分割方法大多从胸部的 PET/CT 图像中进行肿瘤分割，这时得到的肿瘤体积并不是对 MTV_{wb} 的有效估计。而直接将这些方法套用在全身图像上时，由于全身图像远比胸部图像复杂，并不能取得很好的效果，因此需要针对全身 PET/CT 的肿瘤分割方法。

相比于胸部图像，在全身图像中分割肿瘤更加困难。主要的难点如下：

- (1) 胸外肿瘤密度与其周围的正常组织接近，使二者的边界难以区分。
- (2) 在 PET 图像中存在代谢活跃的正常组织，会大量吸收 FDG，如脑、淋巴结、脾脏、肝脏、骨髓、心脏、棕色脂肪，因此可能与同样大量吸收 FDG 的肿瘤混淆。
- (3) FDG 通过肾脏排泄，所以肾脏、输尿管、膀胱可能聚集大量 FDG，影响肿瘤的分割。

1.2 PET/CT 肿瘤分割方法研究现状

1.2.1 在 PET 图像上分割的方法

虽然 PET/CT 是同时拍摄的，但是因为肿瘤主要在 PET 图像中呈现，所以部分方法仅使用 PET 图像进行肿瘤分割，而不使用 CT 图像或仅使其作为辅助。

(1) 基于阈值的方法

阈值法是一种简单、直观的图像分割方法，它通过设定某个阈值，将灰度图中像素值大于该值的像素分为一类，其余像素分为另一类^[5]。在 PET 图像中，像素值反映了细胞的活跃程度，因此肿瘤细胞的像素值通常大于正常细胞。基于这个规律，通过设置合适的阈值，可以将肿瘤从 PET 图像中分割出来。基于阈值的方法主要可以分为以下几类：固定阈值法、适应性阈值法、迭代阈值法。

固定阈值法通过设置一个固定的阈值，对 PET 图像中的所有像素都按照这个阈值进行分类。在许多临床研究中，先将 PET 像素值转换为标准摄取量(standard uptake value, SUV)，然后使用 $SUV = 2.5$ 作为分割肿瘤的阈值^[5]。与之相似的是使用 SUV_{max} 的一个固定比例作为阈值。 SUV_{max} 定义在每个人的完整 PET 图像上，指 PET 图像的所有像素中的最大值。相比于使用固定的 SUV 作为阈值，以 SUV_{max} 的固定比例可以看做是先在不同患者之间进行了标准化，一定程度上减小了个体差异带来的影响。在该固定比例的选择上，最常用的比例是 40%-43%，还有研究使用 50% 也取得了比较不错的效果^[6-8]，Nagel 等人还尝试了 34% 但效果不佳^[9]。固定阈值法有时需要根据不同的 PET 图像调整阈值，其原因是图像的属性、成像设备的类型、图像噪声等因素都会对该阈值的选择产生影响。

适应性阈值法使用数字或物理方法构造模拟的 PET 图像,用来研究从图像性质到某个阈值的映射。也就是说,该阈值是为了适应某种图像性质而选择的。可以用于该目的的图像性质包括源-背景比 (source-to-background ratio, SBR)、平均背景像素值、估计的平均肿瘤像素值和扫描仪的全宽半最大值 (full width half maximum)等^[10]。Nestle 等人使用椭圆形的模拟 PET 图像中不同尺寸的肿瘤拟合出从肿瘤体积到阈值比例的映射^[11],其形式为 $T(\%) = A \times e^{-C \times V}$ 。其中 T 表示阈值占 SUV_{max} 的比例,在每一个 SBR 下,都有两个拟合出的 A 和 C 。 V 表示以立方厘米为单位的肿瘤体积,该体积在实际使用时通过肿瘤在对应的 CT 图像的体积计算获得。Daisne 等人使用 $I_{threshold} = (0.15 \times I_{mean}) + I_{background}$ 计算分割每个肿瘤的阈值^[12]。其中 $I_{threshold}$ 表示该阈值, I_{mean} 表示该肿瘤中的最大像素值 70% 的像素周围的像素的平均值, $I_{background}$ 表示与在该肿瘤相邻的具有最高 SUV 的器官内的平均像素值。该方法考虑到了肿瘤内部的最大像素值可能并不具有统计代表性,而且处于不同器官内的肿瘤边界会受到该器官本身像素值的影响。适应性阈值法的主要缺陷是很难在不同的扫描仪或患者上产生相同或相似的分割结果^[13],这是因为其使用的公式中的参数是针对特定的扫描仪、重构类型、甚至患者身型等因素精调得到的。而且,对于外形复杂的肿瘤,适应性阈值法也难以得到不错的结果。

在应用适应性阈值法分割 PET 肿瘤时,绝大多数依赖于对图像的某种先验估计,如从 CT 中获取肿瘤的大致体积。迭代阈值法^[13]改进了这一不足,它依靠 PET 图像本身迭代产生分割阈值。该方法基于已知体积的球体进行体模测量 (phantom measurement) 得到的各 SBR 下的标定的阈值-体积曲线,从 PET 图像中估计出肿瘤的 SBR,使用标定的 SBR-阈值-体积曲线迭代地计算肿瘤体积。该方法的迭代流程是,先从 PET 图像中获得几条在经典 SBR 下的标定曲线,选择与估计得到的 SBR 曲线最接近的标定曲线。对于该条曲线,存在一个固定阈值 T_1 ,它作用在 PET 图像上,进而根据椭球模型分割出一个肿瘤 V_1 。在迭代时,根据某个体积 V_i 可以产生下一个阈值 T_{i+1} ,如果 T_{i+1} 明显大于 T_i ,那么将 T_{i+1} 再一次应用在 PET 图像上并产生 V_{i+1} 。如果阈值 T_i 和 T_{i+1} 没有明显差别,那么停止迭代,把用 T_i 分割出的 V_i 作为最终分割结果。适应性阈值法需要两个信息,其一是肿瘤的 SBR,这从 PET 图像中容易得到;其二是 SBR-阈值-体积曲线,这由拍摄 PET 图像的成像设备、重构算法和放射示踪剂决定。而且, SBR-阈值-体积曲线严重依赖成

像设备的空间分辨率。此外，该方法假设肿瘤的形状是椭球形，导致肿瘤的形状复杂时的分割精度受到影响。

综上所述，基于阈值的方法的应用相对简单，在肿瘤形状简单时可以取得不错的效果。但是对完整的空间结构考虑很少，当肿瘤形状复杂时会混淆判断。而且，这类方法会受到成像设备等外部因素变化的干扰，健壮性不佳。

(2) 基于随机和学习的方法

基于随机的方法将图像中的像素值分布假设为某种随机变量的概率密度函数，然后通过某种定量统计和优化算法使其可以用于图像分割。基于学习的方法将分割问题用某种模型描述，但是参数未知，需要使用优化算法求解求解。这两种方法具有较强的相似性，因此把它们归为一类。这一类包含混合模型、模糊局部适应性贝叶斯、基于 PET 像素的聚类和分类。

混合模型是一种基于概率分布假设的方法，在此介绍常用的高斯混合模型 (Gaussian mixture model, GMM)。GMM 将 PET 图像的像素值分布与高斯概率密度函数联系起来，将像素值的大小视为随机变量，将像素值在图像中的个数比例视为概率密度。所谓“混合”，指一个 PET 图像可以被看做是多个高斯密度函数的叠加。然后通过优化方法，如最大化期望法，来识别这些概率密度函数，进而达到图像分割的目的。目前已有论文基于 GMM 进行 PET 肿瘤分割^{[14][15]}。前者^[14]首先定义一个初始的区域，然后使用最大化期望法估计多个高斯密度函数中的参数，将该区域内的像素分为目标、背景、不确定三类，基于这些像素产生分割结果。高斯混合模型的优点是可以将像素属于某个类别的可能性用概率量化，具有较强的可解释性。其不足是通常收敛到局部最小值点上，限制了该方法的性能。

后者被称为模糊局部适应性贝叶斯 (fuzzy local adaptive Bayes, FLAB) 法^[15]。FLAB 是一种无监督的方法，它将图像像素分为两个确定类，和这两个确定类之间的一个模糊类。模糊指这类中的像素可以同时以某种程度属于两个确定类。在应用 FLAB 时，需要选择一个感兴趣区域(region of interest, ROI)，因此该方法需要人工参与。FLAB 在肿瘤体积评估上表现出了不错的健壮性和可重复性^{[15][16]}，但是在肿瘤由于部分治疗反应而产生强异质性时会产生错误分割^[17]。而且，FLAB 被设计用于划分两类，当目标类别数大于二时会面临困难。针对这些不足，

有研究提出了一种改进的 FLAB，称为 3-FLAB^[18]。该方法够应对肿瘤的高异质性，而且其划分类别也扩展到了三类，因此在临床上也有一定影响^{[19][20]}。

分类方法通过使用已知标签的数据来划分从图像中提取的特征空间^[21]，其最常用的特征是像素值本身，而且可以从训练数据中学到从数据到标签的映射，进而迁移到无标签的数据上，在合适的条件下可以实现很高的分割水平。但其要求分割目标的变化不宜过大^[22]；数据标签的获取需要人工参与，增加了复杂程度；而且其需要较大的数据量以支持特征提取的稳定性。而聚类方法在数据量小时依然可用，而且可以在目标形态变化较大时发挥作用^[22]。基于像素分类的方法主要包括支持向量机^[23]、人工神经网络^[24]；聚类的方法主要包括 k-近邻^[25]、模糊 C-均值^[26]、谱聚类^[27]、近邻传播^[28]。

(3) 基于区域的方法

基于区域的方法将图像内容的同质性作为决定物体边界的依据。在这类方法中，虽然也考虑到了像素值，但是它更偏向于使用像素值在局部内的分布情况。基于区域的方法主要包括两种：区域生长和基于图的方法。

区域生长法以手动设置的一个种子点开始，以该点周围一个邻域内的像素的平均值和标准差为条件，将与之相邻的像素划分类别。通过设置一个控制同质化的参数集，可以控制一个新的像素在统计上能以何种容忍程度被已经生长的区域吸收。在 PET 图像的分割任务上，当与 $SUV = 2.5$ 和 $SUV_{max} \% = 43\%$ 这两个阈值进行比较时，区域生长法以更小的标准差取得了更稳定的结果^[29]，而且使用区域生长法产生的分割结果可重复性更高。但是，区域生长法严重依赖于种子点的位置和参数的设置。

基于图的方法使用前景和背景种子点作为硬约束，将全局信息与局部成对像素的相似性结合起来产生分割结果^[30]。两种最常用的基于图的方法是图割和随机游走。图割法将图像中的像素构造为图上的节点，将像素间的相似性构造为边。当图构建完成后，通过优化能量函数，使断开边的代价最小化来产生分割结果。当使用局部成对像素的相似性时，图割法被认为可以取得最优的分割结果^[30]，但是它在有噪声的图像上的健壮性不佳^[31]。随机游走法^[32]也被用于 PET 图像分割^[31]。随机游走法对模糊边界和噪声的健壮性更强，因此在 PET 图像中较为适用。

综上所述，基于区域的方法的优点是具有较强的可重复性，不足是大多依赖

初始点的手动指定和参数的选择。

(4) 基于边界的方法

基于边界的方法通过分析 PET 图像中物体的边界而产生分割结果,但是 PET 图像本身的低分辨率和高噪点会给这类方法造成一定困难。基于边界的方法可以进一步分为两类:活动轮廓方法和基于梯度的方法。

经典的活动轮廓方法通过设置一条围绕目标物体的初始曲线,然后使该曲线逐渐运动到目标物体的轮廓上实现分割,这种经典方法称为“Snake”。该曲线通过最小化能量函数来优化。能量函数由内部能量和外部能量组成,其中内部能量描述曲线的平滑程度,外部能量描述曲线与目标边缘的重合程度。在 PET 图像的应用上,Hsu 等人使用 Canny 边缘检测法获得目标的初始轮廓,然后使用 Snake 法用于肝脏分割^[33]。Snake 法要求曲线的初始位置要尽量贴近目标物体,这样外部能量在曲线收敛到目标边缘的过程中才能发挥足够的作用。

水平集法是基于追踪流体在流动的不同阶段之间的界面的一种活动轮廓法。在二维图像分割上,水平集法将 Snake 法对曲线的演化变为对曲面的演化,然后将该曲面与 0 水平面相交时的点的集合作为目标的轮廓线。在涉及界面运动的多个问题上,水平集法都得到了应用,如图像配准和分割。在 PET 图像分割上,Li 等人使用区域生长法作为预处理,然后使用水平集法产生分割结果^[34]。水平集法能较好的处理拓扑变化,而且基于正确的初始化可以分割多个目标。但是能量函数的优化过程较为复杂,而且对初始化的水平要求较高。

基于梯度的方法通过分析图像中像素值的梯度而产生分割结果。每一个像素的梯度使用该像素与其相邻像素的值计算得到。Geets 等人使用分水岭转换和层次聚类分析实现 PET 图像分割^[35]。另外,Werner-Wasik 等人使用了由 MIM software 提出的一种方法,基于手动指定的一个点,计算其半径方向的梯度,然后根据梯度值和边缘连续性给出分割结果^[36]。基于梯度的方法的不足是会受到 PET 图像的模糊性和噪点的干扰。

综上所述,基于边界的方法可以处理一些复杂的边界情况,但是其大多依赖初始设置的准确性,有时需要与其它方法共同使用或人工参与。

1.2.2 在 PET/CT 图像上的分割方法

虽然在 PET 图像上可以进行目标分割,但是其低分辨率和噪声通常会给该过程带来一定干扰。而且,由于其图像模糊,目标的边界具有不确定性。CT 图像的引入可以提供精细的结构信息,为模型的处理提供更多依据,减小分割的不确定性。PET/CT 图像的分割方法主要包括基于活动轮廓模型的方法、基于纹理特征和分类器的方法、基于图的方法、基于概率的方法、基于深度学习的方法。

(1) 基于活动轮廓模型的方法

多值水平集法将边界概念重新定义为多种模态图像的逻辑组合,实现了水平集法的一种推广,其中逻辑组合的权重根据先验信息进行设置^[37]。Khatchadourian 等人首先在切片上选择一系列可能位于肿瘤内部的种子点,然后将误选种子点剔除,使用保留下来的种子点构建肿瘤的初始轮廓,最后使用水平集法产生肿瘤轮廓^[38]。Markel 等人结合了 Jensen-Rényi 散度和水平集,基于灰度直方图的统计发散性优化目标轮廓,从而实现分割^[39]。这类方法虽然具有不错的性能,但是依然依赖初始位置的正确选择。

(2) 基于纹理特征和分类器的方法

这类方法对 PET 和 CT 图像进行定量纹理分析,使用一阶、二阶和高阶纹理特征,结合最近邻和决策树模型进行肿瘤分割^{[40][41]}。优点是使用了多种纹理特征,可以有效的提取图像中的复杂信息。不足是功能型成像与结构型成像的像素点并非是绝对的一对一关系,对这两种信息的结合缺乏统一的标准,而且算法的执行时间较长。

(3) 基于图的方法

Han 等人把分割问题构造成一个图上的马尔科夫随机场的最小化问题,分别构造两个分别描述 PET 和 CT 的子图,同时从 PET 和 CT 图像中分割肿瘤,并加入一个正则化项来惩罚来自二者分割结果的差异^[42]。该马尔科夫随机场使用一个基于图割的方法进行求解。在此基础上, Song 等人在描述 PET 和 CT 的两个子图间添加了适应性的上下文损失,以保证两个模态分割结果的一致性,最后通过基于图割的方法解一个最大流问题得到分割结果^[43]。这些方法的优点是考虑到了在两个模态上的分割结果可能存在差异,因此进行了约束。不足是需要手

动参与，而且依然建立在 PET 和 CT 像素的一对一关系假设上。

Cui 等人把分割问题的求解过程构造为一个迭代的标签传播过程，通过结合 PET 图像中的空间拓扑信息和 CT 图像中的局部像素值变化分割肿瘤^[44]。标签传播的优点是算法逻辑简单，优化速度较快，而且通过利用自身的结构指导标签传播，不需要优化函数。其不足是划分结果不稳定，具有很强的随机性。

(4) 基于概率的方法

Salah 等人通过多视图、多分辨率的隐马尔科夫树和贝叶斯推理来获取各隐状态之间的联合统计依赖，进而实现 PET/CT 肿瘤分割^[45]。Wang 等人基于肿瘤-背景概率模型，用 PET 图像中的像素值下降特征计算一个距离损失，与 CT 图像的背景概率函数结合进行肿瘤分割^[46]。Potesil 等人利用根据 PET 产生的健壮热点和分割产生一个保守的肿瘤分割，然后以此为基础学习 CT 中的肿瘤特征，最后使用联合似然比合并 CT 与 PET 的分割结果^[47]。Lian 等人以信念函数为基础，使用了一种度量学习策略来选择每个像素的特征，以无监督的方式从 PET 图像中分割肿瘤，同时使用 CT 指导该过程，用一个损失项来迭代地约束来自 PET 和 CT 分割结果的冲突^[48]。

(5) 基于深度学习的方法

最近，基于深度学习的方法引起了更多人的关注^[49-54]，也被应用于 PET/CT 图像中不同部位的肿瘤分割。如 Huang 等人提出了一种由特征表示阶段和得分图重构阶段组成的卷积神经网络，用于分割头部和颈部的肿瘤^[55]。Xu 等人提出了 W-Net，将其用于全身骨骼肿瘤的分割^[56]。W-Net 由两个 V-Net^[57]组成。在第一个 V-Net 用于从 CT 图像中分割骨骼；第二个 V-Net 用于从 PET、CT 和骨骼概率图的堆叠中分割肿瘤。Jin 等人提出了 Two-stream-Net 用于食道肿瘤分割^[58]，该模型由三个逐步语义嵌套网络（progressive semantic nested network, PSNN）组成。其中两个 PSNN 分别从 CT 和 PET/CT 中分割肿瘤，第三个 PSNN 从前两者的分割概率图和 CT 的堆叠中分割肿瘤。

在与本文相同的肺癌肿瘤分割任务上，为了结合 PET 和 CT 的特征，Zhong 等人应用了全卷积神经网络从 PET 和 CT 图像中共同分割肿瘤^[59]。Li 等人先从 CT 图像中得到肿瘤的初步概率图，然后从该概率图与 PET 图像的堆叠中提取特征，输出最终概率图^[60]。Zhao 等人将三维的 PET 和 CT 图像分别输入到两个分

离的 V-Net 中,然后将两个 V-Net 的输出做元素级的相加,再使用一个特征融合网络输出分割结果^[61]。Kumar 等人训练了一个 Co-learn-Net 来融合 PET/CT 在胸部切片上的互补信息。该网络产生一系列融合图,用于针对性地改变两个模态在不同类别(肺、肿瘤、纵膈、背景)的融合权重^[62]。

基于深度学习的方法的优点是不需要人工设置特征,而是通过随机初始化的参数提取特征,再以特定的算法进行优化,具有很强的适应性。不足是依赖大量手动标记的数据,获取数据需要消耗大量的人力。

1.3 本文研究内容

本文提出了一种针对全身 PET/CT 图像的肿瘤分割模型。该方法以卷积神经网络为基础,利用了切片的位置先验信息,并以多分支的结构提取和融合模态特征。本文发现以肺的下界为标准,将人体的全身分为上下两个半身,将二者分别处理可以提升模型对样本的针对性,进而提升性能。为此设计了两个解码器,使其各自学习全身复杂特征的一个子集。切片的位置先验信息用于指导模型对人体部位的划分,同时为特征图嵌入隐式的位置信息,以进一步提升模型对疑似区域的判别能力。

本文的贡献可以总结为:

- (1) 利用从 CT 图像中学到的空间关系辅助 PET/CT 肿瘤分割。
- (2) 定义了两个针对人体上下两个半身的解码器,并使用基于位置的门信号控制每张切片在输出的权重。
- (3) 在标准交叉熵中加入了肿瘤级的平衡系数,它以连通域算法和反比例函数为基础,用于缓解不同大小肿瘤在分割损失中权重失衡的问题。

本文与一些现有的方法进行了对比,以说明所提出的方法在全身 PET/CT 肿瘤分割任务上是有效的。而且还应用了生存分析,根据所提方法在测试阶段预测出的肿瘤分割图计算出每个患者的肿瘤体积,然后输入到两个生存分析模型中来评价所提方法在临床应用中的价值。

1.4 本文组织结构

第一章:引言。首先介绍了全身 PET/CT 肿瘤分割方法在临床上的应用价值;

然后对现有的 PET/CT 图像分割方法按照类别进行了简要介绍，并分析了各种方法的优点和不足；其次，提出该研究面临的主要困难和本文思路；最后，对全文结构进行概括。

第二章：卷积神经网络概述。本文所提方法以卷积神经网络为基础，因此在这一章对该方法进行了概述。

第三章：基于 PSMB-Net 的 PET/CT 肿瘤分割。这一章介绍了用于全身 PET/CT 肿瘤分割的模型，对该模型的各个模块进行了详细介绍。

第四章：使用 480 个肺癌患者的 PET/CT 图像验证所提方法，并与其它方法进行了对比。

第五章：总结和展望。这一章列举了所提方法的不足，并提出了改进的思路。

2 卷积神经网络

本章介绍了卷积神经网络的概念，包括卷积神经网络与其方法体系的关系和其发展历程等。随后引出了卷积神经网络的基本构成，包括卷积层、池化层、激活函数、损失函数。最后介绍了卷积神经网络在训练阶段的主要步骤，包括参数初始化、学习率设置、数据预处理和反向传播。

2.1 概念

卷积神经网络属于深度学习模型的一种，而深度学习是机器学习的一个分支。机器学习是通过算法使得机器能从大量数据中学习规律从而对新的样本做决策的过程^[63]。传统的机器学习模型大多依赖人工设计特征，其性能易受到特征选择的影响，如果人工选择的特征与任务需要的特征不匹配，就有可能影响其性能^[64]。为此，就出现了“特征工程”，指将信息用高效的特征表示的方法，同时还能减小原始数据中的不确定因素。特征工程在深度学习出现之前，在计算机视觉领域非常重要。

深度学习模型与传统机器学习模型的一个主要区别是深度学习模型可以自动选择特征，将特征选择与模型的学习过程结合起来，使得特征选择不再是人们需要重点关注的问题。而且随着计算机性能的提升，深度学习在语音识别、自然语言处理、计算机视觉等领域取得了出色的成绩。

卷积神经网络是神经网络的一种形式。神经网络以人类大脑中各神经元的连接为原型，将从数据中提取的特征抽象为节点，将模型对特征的处理过程抽象为节点之间的边，形成一个复杂的网络模型。传统的神经网络由输入层、隐含层、输出层组成。其中输入层就是每一个样本的一系列特征，隐含层是神经网络对样本特征的处理过程，最后从输出层输出。

早期出现的多层感知机就是一种简单的神经网络。最简单的多层感知机由 1 个输入层、1 个隐含层和 1 个输出层组成。所谓“层”，就是并行排列的一系列节点的统称，而节点可以看做是用于存放特征的位置。在输入层，每一个节点中的数字都是样本的一个特征。隐含层中的每一个节点都是上一层所有节点的一种线性组合，也就是用原样本的所有特征组合出一个新的特征。然后从隐含层到输出层，继续重复这个步骤，输出层的每一个节点仍然是隐含层的所有节点的线性组

合。但是输出层具有特定的含义，即要符合我们模型的目标。比如任务是数字 0-9 图像的分类，共有 10 个类，那么一般输出层应设置 10 个节点，以分别对应这 10 个数字。中间的隐含层的节点设置可以根据需要进行调整。

从本质上来说，神经网络是一个函数。在上述这种简单的神经网络的基础上，可以增加隐含层中节点的个数使神经网络变为更加复杂的函数，以完成更困难的任务。常用的方式有增加隐含层的个数和增加每一个隐含层中的节点个数。但是，单纯的增加隐含层的节点个数并不能从根本上增加其复杂程度，因为节点与节点之间只是线性变换，所以从输入层到输出层的多次线性变换后的复杂性增加并不明显。为了显著增强神经网络的复杂性，需要使用非线性函数。人们通常将非线性函数与线性函数交替使用，这样输入层到输出层之间的映射就变得非常复杂，为满足更复杂的任务需要创造了条件。

传统神经网络的每一层节点，都由上一层的所有节点计算而来，这种方式称为“全连接”。虽然这样可以将上一层的信息充分利用，但是上一层的每个节点都有一个模型参数与之对应。这在图像相关的任务上会消耗大量的计算资源和计算时间。比如一张分辨率为 1000×1000 的彩色照片，包含 R、G、B 这 3 个通道，如果使用这种全连接的神经网络对这张照片的内容做分类，那么输入层的节点数等于 $1000 \times 1000 \times 3 = 3,000,000$ （三百万）。对于第一个隐含层的第一个神经元，就需要三百万个参数用于匹配这三百万个输入层的节点，而且隐含层节点的增加带来的模型参数数量的增加是急剧的，这会极大占用计算机的内存和计算时间。这时，“卷积”这种节点之间的连接方式的优点就显得非常突出了。

形象地说，卷积是用一个数组在另一个数组上以某种步长“滑动”，用二者的重合部分计算出下一层的一个节点。其中的一个数组是模型的参数，称为卷积核；另一个数组是输入的特征。相比于全连接的神经网络，卷积神经网络的特点是稀疏交互、参数共享和平移等变^[65]。

稀疏交互指输入图像在同一时刻，只有一小部分与卷积核发生交互而产生特征。也就是说，卷积核的尺寸相对于整个图像来说是很小的。交互的稀疏性造成了提取到的特征是输入图像的某一个小范围内的特征，比如沿某种方向排列的像素、拐角、简单形状。虽然这些独立的小范围特征不能直接用于目标任务，但是因为卷积神经网络有很多层，这些小范围的简单特征会被这些层逐步组合，进而形成可以用于目标任务的复杂特征。稀疏交互大大加速了卷积神经网络提取图像

特征的过程，是卷积神经网络区别于全连接神经网络的主要特征。

参数共享指一个卷积核中的参数在整张图像中所有滑动到的区域所有区域之间，都是相同的。这样使得一个卷积核就可以提取图像的一种特征，使得模型参数量大大减小，有利于模型参数的存储和降低模型的复杂度。

平移等变指卷积前的原图像发生的某种平移变换会在卷积后的特征中保留下来。比如对图像中的所有像素向右平移 3 个像素，然后使用卷积核提取其特征，那么这样操作的结果与先用卷积核提取特征，然后再把特征的所有像素向右平移 3 个像素相等。但需要注意的是，平移变换的等变性不一定对于其它变换适用，比如对图像的缩放或旋转。平移等变使得卷积神经网络对目标所在的位置并不敏感，比如目标是检测一只猫，由于上述的参数共享特性，不论这只猫出现在左上角还是右下角，同一个卷积核提取到的特征都是相同的，都会对其进行正确的检测。因此，平移变换可以看做是参数共享的一个结果，它使得卷积神经网络可以克服位置变化对于目标物体的干扰。

2.2 基本结构

2.2.1 卷积层

(1) 定义

卷积核是组成卷积神经网络的基本单位，对于卷积核概念的介绍需要从信号处理领域说起。在信号处理领域，有两个非常相似的概念：卷积和相关，而相关又分为自相关和互相关。从本质上来说，卷积神经网络中的卷积操作，其实是信号处理领域的互相关。

当变量是一维时，互相关的定义为公式 2.1:

$$(f \star g)(\tau) = \int_{-\infty}^{+\infty} f^*(t)g(\tau + t)dt \quad (2.1)$$

其中 f 和 g 表示两个连续函数（信号）， f^* 表示 f 的复共轭， τ 表示相对位移， \star 表示互相关操作。该式的含义是，将信号 $g(t)$ 平移 τ 个单位后，变为信号 $g(\tau + t)$ ，然后这个信号与 $f^*(t)$ 的积 $f^*(t)g(\tau + t)$ 在 $-\infty$ 到 $+\infty$ 上的积分作为信号 f 和 g 在相对位移 τ 处的互相关值。互相关描述的是两个信号在某一个相对位移处的相似性。如果两个信号在某一位置的互相关值越大，那么这两个信号在这个相对位移处就

越相似。

卷积的定义为公式 2.2:

$$(f * g)(\tau) = \int_{-\infty}^{+\infty} f(t)g(\tau - t)dt \quad (2.2)$$

其中*表示卷积操作，其它各符号的含义与互相关定义中的相同。与互相关的定义比较起来，除信号 f 不取复共轭外，信号 g 关于相对位移 τ 取相反数。形象地说，就是把信号 g 以 τ 为参照点做镜面对称。

自相关的定义是一个函数与自己的互相关，只需要把互相关定义中的一个函数用另一个替代即可。

当变量是二维时，卷积的定义为公式 2.3:

$$(f * g)(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)g(-x + u, -y + v)dxdy \quad (2.3)$$

其中 f 和 g 分别表示两个二维连续函数(信号)， u 和 v 分别表示 f 和 g 在两个维上的相对位移。

一般来说，图像是一个离散的二维函数，适用于离散二维函数的卷积定义为公式 2.4:

$$(f * g)(u, v) = \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} f(x, y)g(-x + u, -y + v) \quad (2.4)$$

其中 f 和 g 是两个二维数组，可以把 f 看做图像，把 g 看做卷积核。那么图像的卷积操作就是先把卷积核 g 旋转 180 度，然后与图像 f 进行重合，二者在每一个对应的离散位置上先相乘再相加，然后得到的数字作为 f 和 g 在位置 (u, v) 的卷积结果。需要注意的是，虽然两个求和符号的下标和上标是无穷，但是实际上卷积核是一个尺寸较小(如 3×3)的数组，那么可以把卷积核看成与图像尺寸相同的一个大数组，但是只有其中 3×3 的区域内的数字是非零的，这样即使是公式的形式上是正负无穷，但是实际上只有很小一个范围参与了运算。

根据上面的介绍得知，图像与卷积核的相关操作与卷积操作的区别就是相关操作不需要将卷积核旋转 180 度。使用卷积核对图像进行处理时，卷积和相关这两种操作可以达到完全相同的结果，因为只需要将卷积操作使用的卷积核翻转即可。在深度学习中，为了应用的方便，实际使用的一般是互相关，因为卷积核中的参数是随着学习过程确定的，本质上都是与图像中对应位置像素的加权和，卷积和互相关的区别只是它们学到的卷积核是否经过翻转而已。

常用的图像除二维外，还有三维图像，如医学图像，通过将一系列沿人体垂直轴拍摄的二维图像堆叠起来而成，可以立体地反应人体内部的情况。用卷积神经网络处理三维图像时，使用的卷积核也是三维的，其过程可以参考从一维到二维的变化。形象地说，是用一个较小的长方体数组沿着三维图像的每一个维度滑动，在其停留的每一个位置上与其邻域内重合的图像部分进行加权求和，作为该位置的卷积结果。

对于多通道的图像，不同通道有不同的含义，如 RGB 图像，三个通道分别表示红、绿、蓝这三种颜色在每一个像素位置的色彩分量。这样的多通道图像在卷积时，卷积核的通道数等于图像的通道数。比如使用一个 5×5 的卷积核对 RGB 图像进行卷积，那么对于图像与卷积核的相对位置 (u, v) ，其卷积结果来自该图像每一个通道内的 (u, v) 邻域内的 $5 \times 5 \times 3 = 125$ 个像素的加权和。

在卷积神经网络中，从原图像中提取的特征有很多种，仍然以输入的是 RGB 图像为例，那么经过第一次卷积之后，每一个卷积核会产生一个通道的特征，因为其每个通道是一个二维数组，所以将其称为特征图。也就是说，一个卷积核将 RGB 图像的这三个通道变为了一个通道。通过增加卷积核的个数，可以从同一张 RGB 图像中提取不同的信息，也就是增加提取到的特征图的通道数。一般来说，一次卷积使用的卷积核在几十个到上百个不等，提取到的特征图也是千变万化。而这些提取到的特征图又被接下来的一系列卷积核再次卷积，经过多次这样的特征提取，卷积神经网络就可以实现从原始图像到目标的映射。虽然这些特征图虽然可以满足目标任务的需要，但是因为都是一些抽象的特征，人们很难用语言去解释它。因此，从整体上来看，卷积神经网络更像是一个“黑盒子”，人们只知道其输入与输出，但是其内部的细节人们无法理解，也就很难做出针对性的改进。不可解释性也因此成为了深度学习模型的一个普遍缺陷。

(2) 步长

图像卷积的一个重要参数就是卷积步长，指卷积核从一个位置滑动到另一个位置后，两个位置之间在各个方向上的距离。以三维医学图像为例，其三个维度分别表示人体的矢状轴、冠状轴、垂直轴。假设该卷积操作的步长是 $(1, 1, 1)$ ，表示卷积核在遍历矢状轴、冠状轴和垂直轴方向时都以 1 个像素的距离移动，这也是最常见的卷积步长。当该卷积操作的步长是 $(1, 2, 3)$ 时，表示卷积核在

矢状轴上以 1 个像素为步长移动，即紧邻上一次滑动的像素；在冠状轴上，在上一次滑动到的像素后隔一个像素；在垂直轴上隔两个像素。卷积操作的步长可以根据任务的需要进行设置。当沿某一个维度的步长大于 1 时，卷积后的图像在该维度上的尺寸会明显缩小。

(3) 奇数尺寸

在卷积核尺寸的选择上，普遍使用奇数，如 3×3 、 5×5 、 7×7 。主要原因有三个：第一，奇数相比于偶数有中心点，对边沿、线条更加敏感，可以更有效地提取这类特征；第二，奇数的卷积核在卷积之前可以从图像的两边做对称的填充；第三，奇数尺寸有中心点，可以方便地从特征反向地定位出关键像素。

(4) 填充

当卷积核与一张图像进行卷积时，卷积得到的图像比原图像在尺寸上有所减小。假设原图像是边长为 m 个像素的正方形图像，卷积核是边长为 k 的正方形数组，且 k 按照上述原因设置为奇数，那么卷积的结果是一个边长为 $m - k + 1$ 的正方形图像，并不等于原图像的边长 m 。这种差异的原因是输入的图像在边界处的每一个像素，由于其外部没有能够与卷积核发生交互的像素，所以卷积核在遍历到这样的像素时会停止继续在该方向的遍历。

在卷积神经网络中，因为有一系列的卷积层，因此特征图的尺寸会因为上述原因而不断减小，最终变为一个像素，导致模型失效。这个现象当 k 取较大值（如 7）时会更加明显。所以需要原图像的边界先进行某种填充，为其卷积导致的尺寸减小留出合适的余地，这样卷积之后的尺寸就能满足任务的需要。

可见填充是卷积操作中重要的一个参数，其方式主要有以下几种：

有效填充 (valid): 虽然名为填充，但实际上这种方式就是不使用填充的极端情况，此时卷积核只能允许与图像中能够将其完全包含的位置进行卷积。图像经过每一次卷积都会减小尺寸，因此会导致上述问题，所以实际上很少会使用这种方式。

相同填充 (same): 相同指用这种方式填充后的图像在卷积后，结果的尺寸与填充前相同。在填充时，使用足够的 0 对原图像的边界进行填充，填充到输出尺寸等于输入尺寸。在这种情况下，只要硬件支持，可以设置任意多的卷积层而保持图像尺寸不会变化。但是这种填充方式导致输入像素中靠近边界的像素会比中

间区域的像素对卷积结果的影响更小,可能会导致边界像素的特征不能很好地提取到。

完全填充 (full): 完全的意思是, 填充足够多的 0 使得原图像中的每个像素在每个方向上恰好被访问 k 次, 也就是说, 卷积核中的每一个值都与每一个原图像素发生了完全的交互。最终输出的图像宽度为 $m + k - 1$, 比原图像尺寸大。

2.2.2 池化

池化^[66]的含义与卷积过程类似, 是使用一个滑窗, 一般设置其步长等于边长 (如尺寸为 2×2 , 步长为 $(2, 2)$), 以使滑窗在滑到的各区域没有重叠, 然后在图像的各维度滑动。使用该滑窗包含的原图像像素的统计量作为这一步的池化结果。常用的池化有: 平均池化, 即计算滑窗内各个像素的平均值; 最大值池化, 即计算滑窗内各个像素的最大值; L2 范数池化, 即计算滑窗内像素值的平方和的平方根; 还有计算滑窗内像素值以各自距离滑窗中心距离为加权的平均值。

卷积神经网络中的池化通常在卷积之后使用, 其目的主要包含以下四点。

第一, 池化可以提高网络对输入的平移不变性。不论使用哪一种池化函数, 当输入像素做出少量的平移时, 池化函数可以使输出近似不变。这是因为每一个池化后的结果, 都与其所在的原图像中滑窗内的邻域像素有某种关系, 这种关系可以在一定的平移内维持不变。比如最大值池化, 如果一个像素是较大范围内的一个孤立的大值, 那么只要它的平移维持在原有的滑窗范围内, 其池化后的结果与平移前是相同的。这种局部的平移不变性可以在关心某个特征是否出现而不关心它出现的具体位置时发挥作用, 比如当判断一张图像中是否包含人脸时, 并不需要知道眼睛所在像素的具体位置, 而只需要知道有一只眼睛在脸的左边, 有一只眼睛在脸的右边即可, 这就允许图像包含一些噪点, 进而提升了模型的健壮性。

第二, 池化可以压缩信息量。通过计算局部的代表性数值, 可以精炼图像的整体结构信息。比如最大值池化, 对滑窗内取最大像素值, 可以更有效地从众多信息中精炼出最有利于目标的信息。再比如平均值池化, 可以综合滑窗内各像素的同时避免最大池化导致的不同区域间的偏差增大带来的不稳定性。

第三, 池化可以减少网络参数。假设池化将原图中的滑窗包含 k 个像素, 那么池化后的特征图的像素量会降低为原来的 $\frac{1}{k}$, 这很大程度上降低了模型对计算机内存的需求量, 而且可以加速后续的卷积过程。

第四，池化可以解决不同尺寸输入图像造成的问题。在很多任务中，池化对于处理不同尺寸的输入有着重要作用。比如需要对不同分辨率的图像进行分类，分类层的节点个数必须是固定的，即类别数。从原图像的不同尺寸到分类层的固定尺寸需要动态调整池化的尺寸和步长来实现^[67]。

虽然池化有这些优点，但是其由于是不可训练的，因此在特征的提取上还存在一定的局限性。为了继承池化的优点，同时使“缩减尺寸”这个操作可以被学习，一些方法中开始使用步长大于 1 的卷积代替池化，如 2×2 ，步长为 (2, 2) 的卷积。这样虽然引入了一些参数，但是使得模型对特征的提取具有更强的任务针对性。

2.2.3 激活函数

激活函数是神经网络的一个重要部分，它是某种一元函数，通过对特征图中每一个元素进行非线性映射，增加模型的表现能力。从神经网络的输入层到输出层，通常会设置很多个激活函数，它们将简单的非线性逐步添加到特征上，最终使得整个模型实现了复杂的非线性。

传统神经网络常使用 Sigmoid 函数^[68]作为激活函数，其定义为公式 2.5:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.5)$$

其定义域为 $(-\infty, +\infty)$ ，值域为 $(0, 1)$ 。其特点是函数图像在 $[-5, 5]$ 这个范围内大致呈 S 形变化，在两端分别接近 0 和 1，变化非常平坦。因此，Sigmoid 函数的优点是可以很容易地抑制绝对值大的噪声。Sigmoid 函数的缺点是，当输入稍微远离坐标原点，函数的导数就变得很小。在神经网络使用反向传播算法优化参数时，经过 Sigmoid 函数后的梯度会变得很小，导致 Sigmoid 函数前的参数更新缓慢，影响模型的收敛过程。而且，Sigmoid 函数要进行指数计算，加大了计算负担。

为了解决 Sigmoid 函数的问题，人们提出了修正线性单元(ReLU)函数^[69]，其定义为公式 2.6:

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (2.6)$$

ReLU 函数的定义模拟了人脑中神经元受到电流刺激的响应。当一个输入电流经过神经元时，如果该电流的强度达到了某个阈值，那么神经元就兴奋，否则就抑制。在 ReLU 函数中，可以把 0 看做这个阈值。当自变量 $x > 0$ 时，ReLU 对 x 做

恒等映射；当 $x \leq 0$ 时，ReLU 不作出任何响应，把函数值映射为 0。

ReLU 函数避免了 Sigmoid 函数的梯度问题，当 $x > 0$ 时，梯度等于 1；当 $x \leq 0$ 时，输入的值等于 0，梯度也等于 0。这增加了特征的稀疏性，使得真正起作用的节点较少，进而增加了模型的泛化能力。但是，这种强制的稀疏性可能屏蔽掉太多有用的特征，导致模型的性能下降。而且，当一个节点处的输出值小于 0，那么该节点接收到的梯度也是 0，可能导致该节点再也不会被任何数据激活，进入”坏死“状态，不但无法提供任何信息，还会白白占用计算资源。

为了解决 ReLU 函数导致的节点”坏死“缺陷，人们提出了 ELU 函数^[70]，其定义为公式 2.7:

$$\text{ELU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{if } x \leq 0 \end{cases} \quad (2.7)$$

ELU 函数在输入为负数的情况下仍输出较小的非零值，这样可以消除节点坏死问题，但是其依然引入了指数运算，会加大计算负担。

PReLU^[71]是 ReLU 的另一种改进型，其定义为公式 2.8:

$$\text{PReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \sigma x, & \text{if } x \leq 0 \end{cases} \quad (2.8)$$

其中 σ 是一个与特征图 x 具有相同通道数的向量，对于 x 的每一个通道， σ 在计算时都有对应的一个分量与该通道内的所有元素相乘。 σ 中的元素可以用反向传播算法优化，一般初始值设置为一个较小的小数。相比于 ELU，PReLU 在负数区间上是线性运算，虽然斜率小，但是不会趋于 0，而且计算起来更加简单。

除了本文列举的这三种，还有很多激活函数，在此不一一列举。在实际应用时，应根据任务的需要选择合适的激活函数。

2.2.4 损失函数

损失函数是神经网络的关键，因为它定量描述了神经网络的输出与训练目标之间的差异，通过对这种差异求导，可以计算出造成这种差异的每一个模型参数的贡献，进而用于模型参数的优化，使整个模型最终满足任务的需要。

不同的任务有不同适应的损失函数，比如将神经网络用于回归时，有平均绝对误差 (mean absolute error, MAE)，定义为公式 2.9:

$$\text{MAE}(y, y') = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (2.9)$$

其中 i 表示某样本在训练的一个批次中的序号， y_i 表示该样本的标签值， y'_i 表示模型对该样本的预测值， n 表示该批次中的样本总数。MAE 度量的是预测与标签之差的绝对值，MAE 对该差值的梯度的绝对值并不会随着该差的变化而变化，在各处都等于 1。因此 MAE 对异常的预测值比较稳健，不会使梯度突然增大。但是当预测值与标签接近时，一个微小的差异可能会使梯度过大，导致其不断在函数值的极小值附近跳动，使得优化出的模型参数不够稳定。

相似的损失还有平均平方误差 (mean square error, MSE)，定义为公式 2.10:

$$MSE(y, y') = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (2.10)$$

其中的符号与 MAE 中的含义相同。MSE 度量的是预测与标签之差的平方，与 MAE 相同的是，二者都不考虑预测偏离标签的方向。不同的是 MSE 对异常值的预测值更加敏感，因为 MSE 对该差的梯度包含该差本身，随着该差绝对值的增加，梯度的绝对值也会增加。因此模型会通过调整参数以使异常值造成的损失更小，而这会牺牲许多正常值，因为这些正常样本的误差比异常值的误差小。MSE 的优点是当预测值与标签值很接近时，产生的梯度很小，这使得参数在收敛时会更加精细地逼近一个解，获得更稳定的优化结果。

将神经网络用于分类时，常用的损失函数是交叉熵损失 (cross entropy, CE)。交叉熵损失来源于香农在信息论中对信息熵的定义。信息熵是对信息的不确定程度的一种定量描述。其定义如公式 2.11:

$$H(X) = - \sum_{i=1}^n p(x_i) \times \ln(p(x_i)) \quad (2.11)$$

其中 X 是一个离散型随机变量，其可能取值为 x_1, x_2, \dots, x_n 。 p 表示事件发生的概率。 $H(X)$ 表示 X 包含的信息熵。信息熵越大，表示 X 发生各事件的不确定性越大。

与信息熵的定义类似，交叉熵损失定义为公式 2.12:

$$CE(y, p) = - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c y_{i,j} \times \ln(p_{i,j}) \quad (2.12)$$

其中 n 表示训练的一个批次中的样本个数， c 表示该任务包含的类别总数。 $y_{i,j}$ 表示第 i 个样本是否属于第 j 类， $y_{i,j} \in \{0,1\}$ ，当属于某类时取 1，不属于某类时取 0。 $p_{i,j}$ 表示模型对第 i 个样本预测属于第 j 类的伪概率，之所以是伪概率，是因为它并不是真正意义上的概率，而是一个取值在区间 $[0,1]$ 的值，用来定量表示样本

属于某一类的可能性。该值常用前面介绍过的 Sigmoid 函数或下面定义的 Softmax 函数获得公式 2.13:

$$p(f_j) = \frac{e^{f_j}}{\sum_{r=1}^c e^{f_r}} \quad (2.13)$$

其中 f_j 表示神经网络的输出层在某一个通道上的某一位置的值, c 表示该任务总共包含的类别数, e 表示自然常数。该公式的含义是, 对于每一个位置, 神经网络会输出多个通道的特征, Softmax 函数先将各个通道位置上的特征值作为 e 的指数, 然后使各通道的值的和等于 1。对于模型输出的较小的 f_j , 其产生的梯度也会较小, 导致优化速度慢, Softmax 函数可以将其放大, 进而加快优化过程; 对于模型输出的较大的 f_j , 其产生的梯度也较大, 导致模型收敛过程不稳定, Softmax 函数可以将其压缩, 进而使收敛过程稳定。

分割任务的本质是对图像中的每一个像素进行分类, 因此分类中常用的交叉熵损失在分割任务上同样适用, 只需要将公式 2.12 在像素的各个可能的位置上计算均值即可。

除交叉熵外, Dice 损失是图像分割中专用的一种损失函数, 它可以缓解交叉熵函数在各类像素的个数差异较大时带来的类别倾向性。其定义为公式 2.14:

$$DiceLoss = \frac{1}{n} \sum_c \left(1 - \frac{2 \sum_{y \in c} y_{true} \cdot y_{pred}}{\sum_{y \in c} y_{true}^2 + \sum_{y \in c} y_{pred}^2} \right) \quad (2.14)$$

其中 $y \in c$ 表示属于类别 c 的像素 y , 共有 n 个类别, y_{true} 表示某像素位置的标签, 且 $y_{true} \in \{0,1\}$; y_{pred} 表示相同位置的预测伪概率, 且 $y_{pred} \in [0,1]$ 。

2.3 训练过程

2.3.1 参数初始化

神经网络是一个非凸优化问题, 参数的初始值很关键。选择合适的初始值, 有利于参数收敛到一个不错的解, 进而获得不错的模型性能。一种常用的初始化是使用在其它任务上训练好的模型拿到自己的任务上, 然后将浅层节点参数保留, 将深层节点随机。如将内容为鸟的图像进行分类, 那么可以使用经过 ImageNet^[72]训练的卷积神经网络中的参数。ImageNet 是一个大型图像数据库, 它包含了超过一千四百万张带标签的图像, 这些图像被分为超过两万一千个类别,

以支持计算机视觉算法的研究。经过这样复杂任务训练的卷积神经网络，其卷积核中的参数可以提取许多复杂的特征，因此常被用于其它网络初始化的参数。由于卷积神经网络在模型的浅层（接近输入层）提取到的特征外观信息更强，如线条、拐角，而在深层（接近输出层）提取到的特征语义信息更强，如猫类、狗类这种人类能够理解的语义类别，人们在使用预训练时往往只保留浅层参数，这是因为浅层参数提取的特征在不同任务之间是相似的，而深层参数提取的特征更有任务针对性，需要根据任务的需要重新对浅层特征进行加工。

虽然使用预训练的模型通常可以提升模型的性能，但是预训练本身加大了目标任务对时间的需要。全部参数都用随机方式初始化是一种更快速的选择。

随机初始化的关键是每一个随机参数服从的概率分布。首先介绍固定值初始化，这是一种特殊的方式，凭借人们的经验进行设置。主要适用于以下网络层：

批量标准化层，其中的 γ 初始化为 1， β 初始化为 0。批量标准化指对批次内的各个样本在各个通道上做标准化，使特征尽量满足标准正态分布，有利于加速学习过程；还可以保持后续网络层的均值和方差与前层相等，减少了两层参数之间的耦合，从而提升健壮性。在标准化之后，使用上述 γ 作为斜率， β 作为偏置对特征进行线性变换，以部分抵消标准化的影响，使该过程是可学习的。

卷积核偏置，通常为 0。但使用 ReLU 作为激活函数时，设置为 0.01，使训练的初期更容易激活。

卷积核中的参数一般使用随机初始化，主要可分为固定方差的随机和缩放方差的随机。

固定方差的随机主要有以高斯分布初始化，使每个参数都服从 $N(0, \sigma^2)$ ；均匀分布初始化，使每个参数在 $[-r, r]$ 以均匀分布初始化，其中 $r = \sqrt{3\sigma^2}$ 。固定方差随机的关键是对方差的选择，如果设置得太小，一会导致卷积后的输出太小，经过多层后会变为 0，二会使基于 Sigmoid 函数的网络损失非线性，因为在 0 附近该函数接近线性变换；如果设置得太大，Sigmoid 函数的梯度接近 0，导致梯度消失。

缩放方差的随机主要有 Xavier 初始化^[73]，设置 $\sigma^2 = \frac{2}{n_{in}+n_{out}}$ 。 n_{in} 表示前一层的卷积核参数个数， n_{out} 表示当前层的卷积核参数个数。Xavier 初始化在激活函数为 Sigmoid 和 tanh 时有很好的表现，但是在 ReLU 时表现不佳。He 初始化^[71]

解决了 Xavier 初始化在 ReLU 上性能不佳问题, 当使用 ReLU 函数时, 设置 $\sigma^2 = \frac{2}{n_{in}}$ 。特别地, 对于 PReLU, 设置 $\sigma^2 = \frac{2}{(1+a^2)n_{in}}$, 其中 a 表示负半轴的斜率。

2.3.2 学习率

学习率用于控制神经网络的参数在每一次迭代中的更新幅度。选择合适的学习率对模型性能的提升很重要。这是因为当学习率过大时, 可能会沿梯度减小方向跳过损失函数的极小值点, 然后循环往复, 使损失函数很难收敛。如果学习率过小, 则会导致每一次参数的更新幅度很小, 收敛速度很慢, 浪费时间。因此在设置神经网络的参数时, 需要尝试多种学习率, 以使得模型达到训练精度与训练速度之间的平衡。一般来说, 会使学习率随着迭代过程的进行而逐渐减小, 使其在一个局部极小值处获得较为稳定的收敛结果。

另外, 因为神经网络的优化是一个非凸问题, 绝大多数的参数会落在极小值点而不是最小值点, 而一旦到达一个极小值点处, 损失函数产生的梯度就会变为 0, 导致无法尝试其它的极小值点, 失去了获得更优解的机会。为了解决这个缺陷, 人们通过学习率的设置, 使参数在优化过程中跳出当前的极小值点。如周期性学习率^[74]、热重启学习率^[75]。它们的核心思想是使学习率在一个大值和小值之间周期性地变化, 其中大值用来使参数跳出极小值点, 小值用来使参数逼近极小值点。

2.3.3 数据预处理

在人工智能领域, 数据会直接影响模型性能的上限。因此, 在应用卷积神经网络时, 需要对数据进行预处理, 使其在模型中能充分发挥作用。

常用的预处理方法是对像素值的拉伸和平移, 以归一化和标准化两种方法为例。所谓归一化, 指的是对一张图像内部的最大像素值等于 1, 最小值等于 0。其定义为公式 2.15:

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.15)$$

其中 Y 表示归一化后的图像, X 表示未归一化的图像, X_{min} 表示 X 中的最小值, X_{max} 表示中的最大值。

归一化可以消除不同图像在量纲上的不同产生的影响。比如, CT 图像的单

位是亨氏单位 (Hounsfield unit, HU), 描述的是对 X 射线的吸收量。空气的 HU 约为-1000, 而人体的大部分软组织 HU 约在[0, 100]范围内, 骨骼一般大于 400。而 PET 图像的标准化单位是标准摄取量 (standard uptake value, SUV), 描述的是代谢活跃程度。空气的 SUV 约为 0, 正常人体组织的 SUV 大部分小于 2.5。当这两种图像同时输入到模型中时, 由于 CT 图像的量纲远大于 PET 图像的量纲, 而模型参数一般是在 0 周围随机分布的较小数, 所以会使 CT 在整个模型的特征中占主导地位, 这不利于 PET 特征的提取。而且, 当使用类似 Sigmoid 的函数时, 会导致很多值分布在定义域的两端, 使得对应的模型参数在反向传播中对损失函数的梯度过小, 不利于参数的优化。

与归一化相似的标准化指将图像中的所有像素值变为服从标准高斯分布, 其定义为公式 2.16:

$$Y = \frac{X - X_{mean}}{X_{std}} \quad (2.16)$$

其中Y表示标准化后的图像, X表示未标准化的图像, X_{mean} 表示X中所有像素的算数平均值, X_{std} 表示X中所有像素的标准差。

除上述两种常用的预处理方法外, 针对特定的数据还需要做额外的处理。以本文使用的 PET/CT 数据为例, 由于 PET 和 CT 的原始图像处于不同的坐标系下, 需要将二者的坐标系统一才能用于模型训练。这种统一坐标系的过程称为配准。另外, PET 原始图像虽然经过配准, 但是其像素值在各个患者之间不具有可比性, 需要将其转换为 SUV 后再使用。

2.3.4 反向传播

卷积神经网络中的参数优化方法是反向传播, 其核心思想是, 在训练阶段的每一次迭代中, 输入的样本在一系列参数的作用下被映射为目标输出, 然后计算损失, 从损失开始逐层求导, 计算损失函数对各个参数的导数, 进而实现参数的逐步优化。下面简单介绍反向传播算法。

设一个卷积神经网络的总层数为 L , 每一次迭代输入 m 个样本 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 。第 l 个卷积层的参数矩阵为 W^l , 偏置为 b^l 。激活函数为 σ 。第 i 个样本在第 l 层的输出记为 $a^{i,l}$ 。损失函数对第 l 层的 $W^l a^{i,l-1} + b^l$ 的梯度为 $\delta^{i,l}$ 。学习率设为 η 。

在每一次迭代时，对于 m 个样本中的一个 x_i ，计算其在第 l 层的输出

$$a^{i,l} = \sigma(W^l a^{i,l-1} + b^l) \quad (2.17)$$

当计算出损失函数的值后，通过损失函数计算出其对输出层的梯度 $\delta^{i,l}$ ，然后计算前面各层的梯度

$$\delta^{i,l} = (W^{l+1})^T \delta^{i,l+1} \sigma'(W^l a^{i,l-1} + b^l) \quad (2.18)$$

然后更新第 l 层的 W^l 和 b^l

$$W^l = W^l - \eta \frac{1}{m} \sum_{i=1}^m \delta^{i,l} (a^{i,l-1})^T \quad (2.19)$$

$$b^l = b^l - \eta \frac{1}{m} \sum_{i=1}^m \delta^{i,l} \quad (2.20)$$

重复执行输入样本、计算梯度、更新参数，直到满足停止条件为止。

需要注意的是，本文只介绍反向传播中最常见的一般形式，在卷积神经网络中的实际情况会更加复杂，需要根据具体情况计算，因此不作展开说明。

2.4 本章总结

本章首先介绍了卷积神经网络的概念和它与机器学习、深度学习、神经网络的关系，它具有稀疏交互、参数共享和平移等变的特点。然后介绍了卷积神经网络主要由卷积层、池化层、激活函数和损失函数组成，描述了各部分的原理与核心概念。最后介绍了卷积神经网络在训练过程中的参数初始化、学习率设置、数据预处理和反向传播算法。

3 基于 PSMB-Net 的 PET/CT 肿瘤分割

本章提出了一种位置监督的多分支神经网络 (position supervised multi-branch network, PSMB-Net) 来同时学习 CT 切片的位置特征和肿瘤分割相关特征。与 MVP-Net^[76]相似, 位置监督的作用是促使模型提取到的 CT 特征是位置敏感的, 进而使其具有更强的位置可分能力, 最终提升模型的肿瘤分割性能。每张 CT 切片的相对位置使用手动标记的方式在 PSMB-Net 的训练阶段开始前产生, 用于指导模型对位置敏感特征的学习过程。基于每张切片的相对位置, 本文设置了两个解码器 (decoder), 分别用于人体的上半部分和下半部分, 这样它们各自会有更强的样本针对性。在标准交叉熵中加入了肿瘤级的平衡系数, 缓解了数据分布的不平衡问题。PSMB-Net 的损失函数包含 3 项, 分别是位置损失 \mathcal{L}_{pos} 、分割损失 \mathcal{L}_{seg} 和正则化损失 \mathcal{L}_{regu} 。

3.1 模型结构

PSMB-Net 由 4 个部分组成, 包含(A)CT 切片的位置监督, (B)多尺度 CT 特征融合, (C)门信号生成, (D)模态特征融合与肿瘤分割, 如图 3.1。

3.1.1 CT 切片的位置监督

PSMB-Net 的第一个分支, 如图 3.1 部分 A, 将三维 CT 图像输入到模型中, 预测其包含的每张切片的相对高度值。因为二维卷积操作以切片为单位提取特征, 而不考虑切片间原有的位置关系, 所以将表示垂直轴的维度按照其固有顺序转换到表示批次的维度以满足该操作的要求。为了更好地利用切片沿垂直轴的空间信息, 每张切片与各自相邻的上下两张切片在通道维度上堆叠, 使其作为额外的特征。经过该操作后, 需要 3 步来实现从 CT 切片到相对位置的映射。

首先, CT 切片经过一系列的卷积层和 PReLU 函数。对于一个三维卷积核, 其卷积过程可以用公式 3.1 表示:

$$W \otimes X_{(x,y,z)} = \sum_i \sum_j \sum_k W_{(i,j,k)} \times X_{(x-i,y-j,k-z)} \quad (3.1)$$

其中 \otimes 表示卷积操作。 $X_{(x,y,z)}$ 表示在一个三维图像 X 中, 处于坐标 (x, y, z) 上的值。其中 x, y, z 分别沿矢状轴、冠状轴、垂直轴变化。对于二维卷积, $z = 1$ 。

W 是一个维数与 X 相等的数组，其中的每个值以随机的方式初始化，然后用反向传播法进行优化。

然后，一个全局平均层将每一张切片的每一个通道平均为一个值。这时，每一张切片在模型中表示矢状轴和冠状轴的维度长度都为 1，因此可以等效地将这两个轴移除，这时四维特征就变为二维数组。该数组的每一行表示一张切片的一系列特征，每一列表示来自不同切片的同一种特征。最后，对每张切片的一系列特征，使用全连接层将其映射为一个值，即这张切片所对应的相对高度。在训练阶段的每次迭代中，用这些预测出的值与位置标签计算出最小平方损失，进而优化该分支。

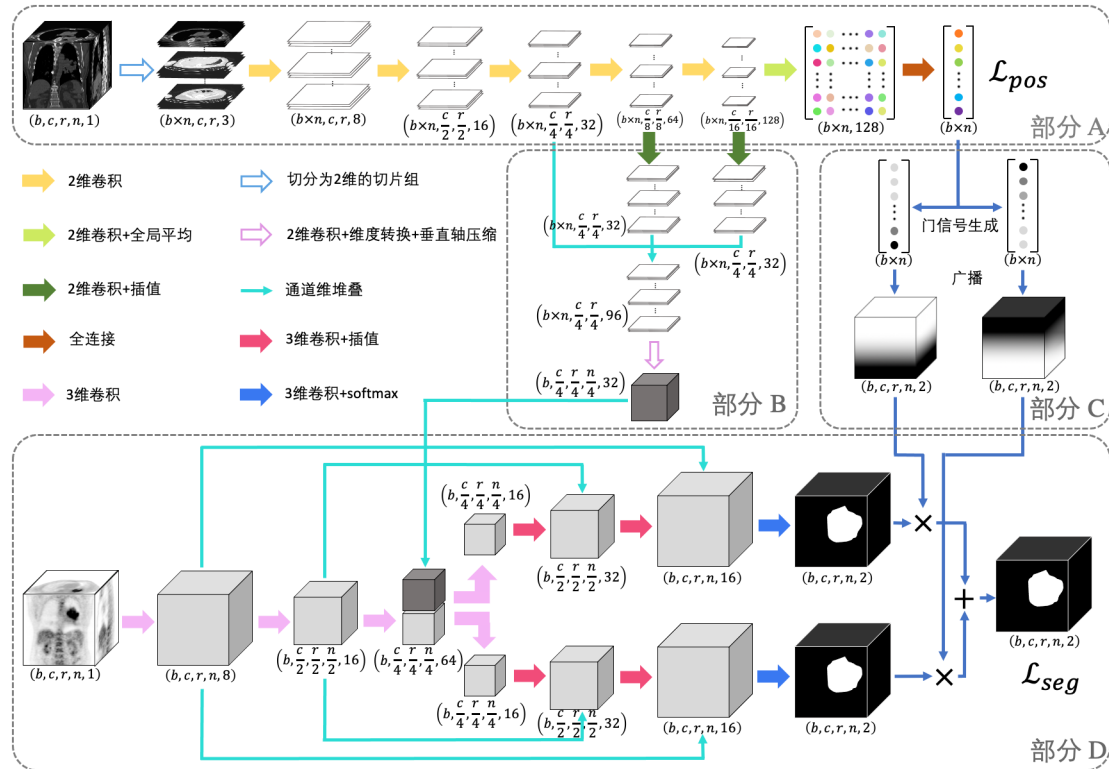


图 3.1 PSMB-Net 的整体结构，包括位置监督分支（部分 A）、CT 特征融合分支（部分 B）、门信号生成分支（部分 C）和模态特征融合与肿瘤分割分支（部分 D）。

3.1.2 CT 特征融合

PSMB-Net 的第二个分支，如图 3.1 的部分 B，描述了从 CT 的二维特征到三维特征的变换过程，该过程的目的是将两个模态的特征进行融合。随着网络的加深，提取到的特征图会有更强的语义信息，但是其代价是空间分辨率会迅速减小，并且很多细节信息会丢失。受到全景特征金字塔网络^[77]的启发，将全局平均层之

前的连续的 3 个尺寸的 CT 特征图融合到一起。具体来说，这 3 个连续尺寸的特征图的通道数分别为 32、64、128，将后两个特征图的通道先用 1×1 卷积压缩到 32，然后将这两个特征图的空间尺寸以线性插值的方式扩大到与第一个特征图相等。将这 3 个特征图在通道维上堆叠起来，形成一个 96 通道的特征图，然后使用 1×1 卷积将其通道数压缩为 32，以减小计算量。最后把切片从批次维度中按照其原有顺序恢复到表示垂直轴的维度中，这样就将 CT 特征转换到了与将要融合的 PET 特征相同的形状。

3.1.3 门信号生成

将每张 CT 切片的相对位置转换为其对应的门信号，然后该门信号作用于每个 BPS-decoder 上，使每个 BPS-decoder 学习一个特定的特征子集。将人体的全身分为两个半身，划分依据是肺部的下界。用于人体上半部分的门信号记为 G_{upper} ，用于人体下半部分的门信号记为 G_{lower} 。

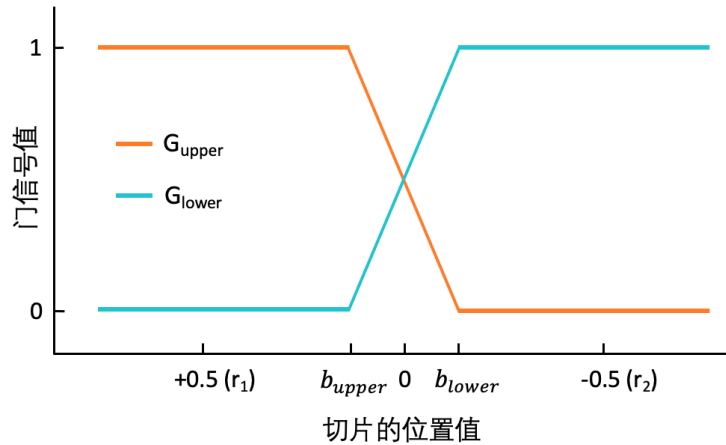


图 3.2 两种门信号的值随切片的相对位置的变化趋势

为了增加模型的健壮性，我们遵循 Zhu 等人提出的软距离门信号的形式^[78]，定义了位置门信号。对于第 i 张切片 x_i ，它的两个门信号 G_{upper} 和 G_{lower} 定义为公式 3.2 和公式 3.3:

$$G_{upper}(x_i) = \begin{cases} \frac{x_i^h - b_{lower}}{b_{upper} - b_{lower}}, & \text{if } b_{upper} > x_i^h > b_{lower} \\ 1, & \text{if } x_i^h \geq b_{upper} \\ 0, & \text{if } x_i^h \leq b_{lower} \end{cases} \quad (3.2)$$

$$G_{lower}(x_i) = 1 - G_{upper}(x_i) \quad (3.3)$$

其中 x_i^h 表示切片 x_i 对应的相对位置。 b_{upper} 和 b_{lower} 表示门信号随相对位置呈线性变化区域的上界和下界，如图 3.2。在本文中，设置 $b_{upper} = 0.05$ ， $b_{lower} = -0.05$ 。在这一设置下，绝大多数患者肺部下界所在的切片的相对位置会处于区间 $[b_{lower}, b_{upper}]$ 内，为不同患者之间的差异留出了一定的缓冲空间。不在该范围内的患者也会非常接近两个边界值，因此不会对模型造成严重影响。

产生的 G_{upper} 和 G_{lower} 是两个长度为 $b \times n$ 的向量，与恢复 CT 特征的形状的过程相似，先按照原有切片顺序从批次维度中提取原图像中的连续切片，使其形状变为 (b, n) ，再用广播的方式把它们的形状扩展到 $(b, c, r, n, 2)$ ，用于做元素级的门抑制 (gating)。广播意味着首先在 G_{upper} 和 G_{lower} 中添加长度等于 1 的维度，此时它们的形状是 $(b, 1, 1, n, 1)$ 。然后将这个 5 维数组在矢状轴维度、冠状轴维度、通道维度上分别复制 c 、 r 、2 次。对于每张切片来说，它所有空间位置的所有通道的值都是相等的。

在训练阶段，两个位置门信号 G_{upper} 和 G_{lower} 是用位置标签计算的。在测试阶段，它们是用 PSMB-Net 的部分 A 的预测结果计算的，因此在测试阶段，不需要位置标签。

3.1.4 模态特征融合与肿瘤分割

PSMB-Net 的第 4 个分支，如图 3.1 的部分 D，描述了 CT 与 PET 两个模态的特征融合与肿瘤分割过程。考虑到肿瘤在立体空间中的特征变化多样，该分支使用 3 维卷积核来提取每个体素邻域内丰富的空间信息。

该分支由一个共享的编码器和两个解码器 (BPS-decoder) 组成。共享的编码器从原始 PET 图像中提取普遍性特征，两个分离的 BPS-decoder 从 PET/CT 融合特征图中各自提取针对性特征。为了使模型的输出与输入有相同的空间尺寸，在 BPS-decoder 中使用线性插值来将特征图的空间分辨率恢复到与输入的图像相同。为了保存更多的细节信息，对每个 BPS-decoder 使用了跳跃连接，其含义将在编码器内部产生的中间结果，以通道堆叠的方式与在每个解码器中对应尺寸的刚经过插值的特征图融合起来。BPS-decoder 中的所有操作都是相同的。在每一个 BPS-decoder 的末端，使用尺寸为 $1 \times 1 \times 1$ 的卷积核将它的输出结果的通道数映射为类别数。然后一个体素级的 Softmax 函数将每一个体素位置上的两个通道映射

为该体素位置属于两个类别的伪概率，Softmax 如公式 2.13:

$$p(f_i) = \frac{e^{f_i}}{\sum_{r=1}^R e^{f_r}} \quad (2.13)$$

每个 BPS-decoder 都会产生各自的 $p(f_i)$ ，而且每一个 BPS-decoder 都对应各自的门信号 (G_{upper} 或 G_{lower})。将每个 BPS-decoder 与对应的经过广播的门信号做元素级的乘法，然后将这两个积做元素级的加法，就得到了最终的概率图。

PSMB-Net 的部分 A 和部分 D 的操作细节如表 3.1。

表 3.1 PSMB-Net 的部分 A 和部分 D 的输入特征形状与对应操作

输入特征的形状	操作	输入特征的形状	操作
	部分 A		部分 D (共享的编码器)
(b,r,c,n,1)	切分为三维切片组	(b,r,c,n,1)	(3×3×3 普通卷积+PReLU)×2
(b×n,r,c,3)	(3×3 普通卷积+PReLU)×2	(b,r,c,n,8)	3×3×3 下采样+PReLU
(b×n,r,c,8)	3×3 下采样+PReLU	(b, $\frac{r}{2}, \frac{c}{2}, \frac{n}{2}, 16$)	(3×3×3 普通卷积+PReLU)×2
(b×n, $\frac{r}{2}, \frac{c}{2}, 16$)	(3×3 普通卷积+PReLU)×2	(b, $\frac{r}{2}, \frac{c}{2}, \frac{n}{2}, 16$)	3×3×3 下采样+PReLU
(b×n, $\frac{r}{2}, \frac{c}{2}, 16$)	3×3 下采样+PReLU	(b, $\frac{r}{4}, \frac{c}{4}, \frac{n}{4}, 32$)	(3×3×3 普通卷积+PReLU)×2
(b×n, $\frac{r}{4}, \frac{c}{4}, 32$)	(3×3 普通卷积+PReLU)×2	(b, $\frac{r}{4}, \frac{c}{4}, \frac{n}{4}, 32+32$)	(1×1×1 普通卷积+PReLU)×2
(b×n, $\frac{r}{4}, \frac{c}{4}, 32$)	3×3 下采样+PReLU		部分 D (每一个解码器)
(b×n, $\frac{r}{8}, \frac{c}{8}, 64$)	(3×3 普通卷积+PReLU)×2	(b, $\frac{r}{4}, \frac{c}{4}, \frac{n}{4}, 64$)	1×1×1 普通卷积+PReLU
(b×n, $\frac{r}{8}, \frac{c}{8}, 64$)	3×3 下采样+PReLU	(b, $\frac{r}{4}, \frac{c}{4}, \frac{n}{4}, 16$)	插值
(b×n, $\frac{r}{16}, \frac{c}{16}, 128$)	(3×3 普通卷积+PReLU)×2	(b, $\frac{r}{2}, \frac{c}{2}, \frac{n}{2}, 16+16$)	(3×3×3 普通卷积+PReLU)×2
(b×n, $\frac{r}{16}, \frac{c}{16}, 128$)	2D 平均池化	(b, $\frac{r}{2}, \frac{c}{2}, \frac{n}{2}, 16$)	1×1×1 普通卷积+PReLU
(b×n,1,1,128)	1×1 普通卷积+PReLU	(b, $\frac{r}{2}, \frac{c}{2}, \frac{n}{2}, 8$)	插值
(b×n,1,1,128)	1×1 普通卷积+PReLU	(b,r,c,n,8+8)	(3×3×3 普通卷积+PReLU)×2
(b×n,1,1,32)	1×1 普通卷积+PReLU	(b,r,c,n,8)	1×1×1 普通卷积
(b×n,1,1,1)	位置监督	(b,r,c,n,2)	Softmax+位置门控

3.2 相对位置标签的计算

本文在每个患者的三维 CT 图像上选择了两张切片，作为计算每张切片的位置的参照，如图 3.3 (a)，该步骤只需要在模型的训练阶段前进行。其中一张切片选择在肱骨大结节所在的高度上，另外一张切片选择在腰椎的最下端。这两张切片在全身 CT 切片中的序号分别记为 r_1 和 r_2 。二者的平均值记为 S_{zero} ，而且切片序号为 S_{zero} 时对应的相对位置等于 0。对于一张切片，如果它在 CT 图像中的位置高于 S_{zero} ，那么它的相对位置大于 0；如果它在 CT 图像中的位置低于 S_{zero} ，那么它的相对位置小于 0。需要注意的是， S_{zero} 并不需要对应确定的一张切片，

即它的取值可以是小数,因为它只是用于计算每张切片相对于它的位置的一个参照。

计算出的切片的相对位置标签是真实空间中每张切片相对于 S_{zero} 的距离与两张参照切片间的距离之比,因此在计算时还计算时需要考虑 CT 切片的层间距,如公式 3.4:

$$d_i = \frac{(S_{zero} - i) \times T}{(r_2 - r_1) \times T} = \frac{r_1 + r_2 - i}{r_2 - r_1} = \frac{r_1 + r_2 - 2 \times i}{2 \times (r_2 - r_1)} \quad (3.4)$$

其中 i 表示某张切片在它所在的全身 CT 图像中的序号,且每个人的 CT 图像都沿从头到脚的方向递增序号,第一张切片的序号值等于 1。 d_i 表示该切片的高度值, T 表示这个人 CT 切片的层间距,即两张 CT 切片扫描厚度中点间的距离。可以发现,虽然在公式的定义中考虑了层间距的影响,但是在实际计算过程并不需要,因为它被抵消了,这说明只使用两张参照切片的序号计算出的相对位置在不同层间距的人之间具有可比性。而且,每个人的所有切片对应的相对位置是随着从头到脚的方向严格线性递减的,如图 3.3 (b)。

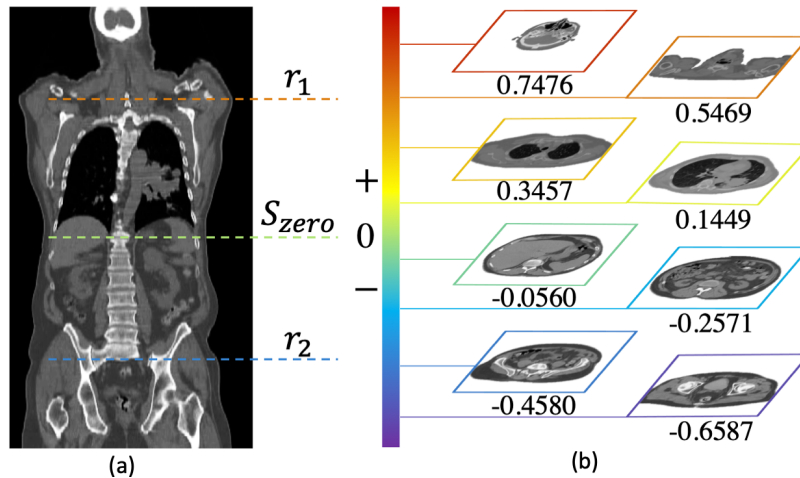


图 3.3 相对位置标签计算示意图。(a)参照切片与零高度的位置;(b)该样本对应的 8 张等距切片对应的高度值。

3.3 肿瘤级的平衡系数

在 PET/CT 数据上不同肿瘤的体素个数不平衡,会导致训练时较小肿瘤无法在迭代中产生足够的影响,进而在测试时容易被忽略。本文在标准交叉熵中加入了肿瘤级的平衡系数来缓解这个问题,其方式是将每一个肿瘤的体素个数映射为该肿瘤中的每一个体素在分割损失中的权重,如公式 3.5:

$$f(T_i) = \frac{a}{T_i^b} \quad (3.5)$$

其中 T_i 是第 i 个肿瘤的体素个数， $f(T_i)$ 是这个肿瘤中每一个体素在损失函数中的权重，该权重以在肿瘤标签图像中的像素值的形式存在。 a 是一个固定参数，它用于将所有肿瘤类的权重缩放到合适的范围内，且 $a \in (0, +\infty)$ 以保证在平衡之后的梯度的方向不会变化。 b 是一个固定参数，用于平衡拥有不同体素个数的肿瘤的权重，且 $b \in (0, +\infty)$ 以抑制大肿瘤的贡献同时放大小肿瘤的贡献。但如果 $b > 1$ ，小肿瘤整体的权重会大于大肿瘤整体的权重，这可能会导致新的不平衡问题。

3.4 损失函数

PSMB-Net 的损失函数如公式 3.6，它包含了位置损失 \mathcal{L}_{pos} （平均平方）、分割损失 \mathcal{L}_{seg} （交叉熵）和正则化损失 \mathcal{L}_{regu} （ ℓ_2 范数）。

$$\begin{aligned} \mathcal{L}_{total} &= \alpha \times \mathcal{L}_{pos} + \beta \times \mathcal{L}_{seg} + \gamma \times \mathcal{L}_{regu} \\ &= \alpha \times \frac{1}{bn} \sum_b \sum_{i=1}^n (d'_i - d_i)^2 + \gamma \times \frac{\|\omega\|_2^2}{2} \\ &\quad + \beta \times \frac{1}{bcrn} \sum_b \sum_{i=1}^c \sum_{j=1}^r \sum_{k=1}^n \sum_{m=1}^M -\eta_{i,j,k}^m \times y_{i,j,k}^m \times \ln(y'_{i,j,k}^m) \end{aligned} \quad (3.6)$$

其中 b 表示每一个批次中的三维样本块（patch）的数量。 n 表示每个样本块中包含的切片张数， c 表示每张切片的像素列数， r 表示每张切片的像素行数。 d_i 是第 i 张切片相对位置的标签， d'_i 是模型对第 i 张切片相对位置的预测。 ω 是 PSMB-Net 中除 PReLU 中的参数外的所有可训练的参数组成的向量。 $y_{i,j,k}^m$ 和 $y'_{i,j,k}^m$ 分别表示 i, j, k 处的体素属于第 m 类的概率，且 $y_{i,j,k}^m \in \{0,1\}$ 为医生提供的标签， $y'_{i,j,k}^m \in [0,1]$ 为模型的预测。 M 表示总类别数。 $\eta_{i,j,k}^m$ 用于平衡在 i, j, k 处的第 m 个类别的梯度。 α, β 和 γ 是三个损失项的权重。

3.5 本章总结

本章提出了一种用于全身 PET/CT 肿瘤分割的方法，称为 PSMB-Net。PSMB-Net 由四个分支组成，一个分支用于从 CT 切片中提取位置敏感特征，产生的每

张切片的相对位置被另一个分支映射为门控信号；一个分支用于融合多尺度的 CT 特征，实现 PET/CT 的模态融合；最后一个分支包含一个编码器和两个解码器，用于学习肿瘤分割特征。两个解码器的输出以门信号作为权重求和作为最终分割结果。另外，在交叉熵损失中加入了一个肿瘤级平衡系数用于提升模型对小肿瘤的关注程度。

4 实验

本章使用 480 个非小细胞肺癌（non-small cell lung cancer, NSCLC）患者的 PET/CT 图像来验证所提方法的性能。在全身图像和胸部图像上与 7 个现有 PET/CT 分割方法进行了比较。在全身图像上进行了消融实验以验证核心模块的有效性。使用图像分割指标和生存分析指标来量化模型的性能。

4.1 数据

表 4.1 数据的特征统计

总数:	480 (100%)	组织学亚型, 个数 (%)	
生存时间(天), 中位数(四分区间):	652.5 (264.5-1821)	腺癌	244 (50.8%)
年龄:均值, 标准差(年)	68.4 (10.7)	鳞状细胞癌	144 (30.0%)
年龄:阶段, 个数 (%)		大细胞癌	21 (4.4%)
34 或更小	1 (0.2%)	非特指型 NSCLC	61 (12.7%)
35-44	6 (1.3%)	其它	10 (2.1%)
45-54	51 (10.6%)	吸烟史, 个数 (%)	
55-64	120 (25.0%)	从未	39 (8.1%)
65 或更大	302 (62.9%)	目前在吸	170 (35.4%)
肤色, 个数 (%)		曾吸	271 (56.5%)
白	186 (38.8%)	临床 TNM 分期, 个数 (%)	
黑	280 (58.3%)	I	116 (24.2%)
其它	14 (2.9%)	II	59 (12.3%)
治疗类型, 个数 (%)		III	153 (31.9%)
已经手术	156 (32.5%)	IV	152 (31.7%)
未经手术	273 (56.9%)	MTV_{wb} 级别, 个数 (%)	
未治疗	51 (10.6%)	I	131 (27.3%)
性别, 个数 (%)		II	129 (26.9%)
女	267 (55.6%)	III	112 (23.3%)
男	213 (44.4%)	IV	108 (22.5%)

本文实验的数据为 480 个肺癌患者的全身 PET/CT 图像，该数据中的患者身份信息已经被抹除，且数据在该项研究中的使用受到了许可。这些 PET/CT 图像

的拍摄时间为从 2004 年到 2014 年。其中 352 人的 PET/CT 图像使用 Reveal HD 成像仪在 2012 年 3 月 15 日之前拍摄, 128 人的 PET/CT 图像使用 Siemens mCT 成像仪在 2012 年 3 月 15 日或其之后拍摄。每人的 PET/CT 切片经过配准, 且使用的分辨率是 256×256 像素。

原始的 PET 图像中的体素值被转为 SUV, 以使其在不同人之间具有可比性。最终使用的每个人的 PET/CT 切片数量不一定相等, 其范围是 189 到 543, 且每人体内的肿瘤个数从 1 到 57。用于分割的肿瘤标签由 2 位经验丰富的放射科专家使用 MIMvista 5.1.2 勾画得到。这些患者的统计信息如表 4.1。

4.2 实现细节

在全身 PET/CT 图像上, 包含肿瘤的切片张数远小于不包含肿瘤的切片张数, 所以在对训练数据采样时, 对每次提取的不包含肿瘤的样本块有 80% 的丢弃概率, 以进一步平衡两类差距, 且该操作在所有方法上都应用。在公式 3.6 中, 设置了 $\alpha = 0.1$, $\beta = 1$, $\gamma = 5 \times 10^{-6}$; $\eta_{i,j,k}^m$ 是通过公式 3.5 在训练阶段开始前计算的。在公式 3.5 中, 设置了 $a = 10000$, $b = \frac{3}{4}$ 。所有方法输出的概率图都使用 DenseCRF^[79]进行优化。

对于所有方法, 使用 5 折交叉测试。480 个患者的顺序被随机打乱, 而且对于每一次训练-测试的循环, 96 个患者用于测试, 其它患者用于训练。对于所有可训练的方法, 使用了数据增广, 沿着冠状轴以 50% 的概率随机翻转。每一个用于训练的样本块由连续的沿垂直轴分布的切片组成。当切片不能填满一个样本块时, 使用覆盖填充来解决。

实验环境是 Python 3.7.3 和 Tensorflow 1.14.0。图形处理器是英伟达 Tesla P100, 包含 16GB 的运行内存。在训练阶段, 设置的学习率是 $10^{-4} \times 0.99^m$, 其中的 m 从 0 开始, 且每经过 500 次迭代后加 1。总共的迭代次数是 10^4 。在激活函数 PReLU 中, 可训练的参数初始为 0.25。在本文中定义样本块是包含连续切片的图像体。在全身的比较中, PSMB-Net、V-Net、W-Net、Two-stream-Net 的样本块包含 64 张切片, 但是受到 GPU 内存的限制 Co-learn-Net 的样本块包含 32 张切片。在全身比较中, 上述所有方法的样本块都包含 32 张切片。设置上述方法的批次大小为 1, 即训练的每次迭代时输入 1 个样本块。

4.3 其它方法的实现

本文把所提方法与三个现有的 PET/CT 分割方法: W-Net^[56], Two-stream-Net^[58], Co-learn-Net^[62]; 两个通用的医学图像分割方法: V-Net^[57]和 DeepMedic^[81]; 两个基于阈值的方法^[82]: 42% SUV_{max} 和 50% SUV_{max} , 进行了对比。这些方法的实现细节如下。

W-Net 中的两个 V-Net 使用了相同的设置, 其通道数在 3 个特征尺寸上都是 (16, 32, 64), 第一个 V-Net 以 CT 图像作为输入, 以肺概率图作为输入第二个 V-Net 以 PET、CT、肺概率图的堆叠作为输入, 以肿瘤概率图作为输出。

Two-stream-Net 包含三个称为 PSNN 的子网络, 且它们是分别训练的。因此在训练时先训练整个流程中的前两个 PSNN, 忽视第三个, 而且前两个 PSNN 在训练时的梯度不会相互影响。经过 80000 次迭代后, 固定了前两个 PSNN 中的参数, 开始训练第三个。整个训练过程的总迭代次数是 120000。

在提出 Co-learn-Net 的论文中, 在训练前额外产生了肺部和纵膈的标签。因此我们在所用的实验数据上产生了肺部和纵膈的标签。根据该论文的代码^[80]重现了该模型, 以适用所使用的三维数据和其它操作, 例如样本块的随机丢弃和后处理。在通道的设置上, 在普通卷积层上设置为 32, 在联合学习层上设置了 64 以适应特征融合。

V-Net 设计用于一个模态, 但是在本文任务中有两个模态。将 PET 与 CT 图像在通道维度上进行堆叠以解决这个问题。与 W-Net 中的每个 V-Net 相同, 在不同特征尺寸上的通道数设置为 (16, 32, 64)。

DeepMedic 是一个三维卷积神经网络, 它由多个支路组成。它的初始用途是多模态 MRI 图像的多尺度处理, 而且很容易将它的用途扩展到 PET/CT 肿瘤分割任务上。实验使用了该论文的公开代码^[83], 而且使用了其初始设置, 但是受到图形处理器内存的限制, 将其中的批次大小从 10 降低到 8。读者可以通过获取阅读论文和代码来获取更详细的信息。

SUV_{max} 指将每个患者的 PET 图像标准化为 SUV 后, 所有体素值中的最大值。以该最大值的比例作为基础的阈值法是一种从 PET 图像中分割肿瘤的传统方法。两个常用的阈值分别是 42% SUV_{max} 和 50% SUV_{max} 。将每个患者的 SUV 体素值从大到小排序, 然后选择最大的 10 个值的平均值作为该患者的 SUV_{max} , 以提

高健壮性。

一个现有的相对位置的计算方法^[84]可以与手动产生的相对位置有相同的用途，即在 PSMB-Net 中用于位置监督。为了展示手动标记的重要性，将使用手动产生的相对位置与该方法产生的相对位置分别作为 PSMB-Net 中位置监督的标签，然后比较模型的肿瘤分割性能。

本文使用与提出该方法的论文介绍的相同设置。在训练的每次迭代时，随机选取了 12 个患者的全身 CT 图像，然后对于每一个患者选取 8 张等距的切片。对于每一个全身 CT 图像，选取的等距切片之间的序号差是在区间[1,20]的整数。使用与 PSMB-Net 中相似的结构将每一张 CT 切片映射为一个相对位置。基于每张切片的相对位置，一个顺序约束项和一个距离约束项组成的损失函数用来优化该模型。使用线性变换 $y = -\frac{1}{40}x + 0.25$ 将该模型的输出值映射到与手动标记的方式产生的相对位置大致相等的区间内，以求更方便和公平的比较。

4.4 结果

4.4.1 指标

本文使用 Dice、Sensitivity、Precision 这三个指标来衡量所有方法的性能。其中 Dice 是预测与标签的交集与它们的并集之比，是一个综合性的指标；Sensitivity 也称灵敏度，反映了对肿瘤预测的全面程度；Precision 也称准确度，反映了对肿瘤预测的准确程度。它们的定义分别如公式 3.7、3.8、3.9：

$$Dice = \frac{2 \times |V_P \cap V_G|}{|V_P| + |V_G|} \quad (3.7)$$

$$Sensitivity = \frac{|V_P \cap V_G|}{|V_G|} \quad (3.8)$$

$$Precision = \frac{|V_P \cap V_G|}{|V_P|} \quad (3.9)$$

其中 V_P 表示预测的肿瘤掩膜， V_G 表示标签的肿瘤掩膜。 $|V|$ 表示 V 中的非零体素值的个数。需要注意的是，只有肿瘤类的分割图像才用于计算上述指标。如果某个患者的 $|V_P|$ 和 $|V_G|$ 都是 0，则表示实际上该患者没有肿瘤，模型也没有分割出肿瘤，这时对应的指标定义为 1。如果只有其中一个是 0，那么计算出的指标定义为 0。

4.4.2 比较和分析

每位患者的指标是分别计算的，然后在 480 位患者上取平均值。

首先与现有方法在全身 PET/CT 图像上进行了比较，所有方法在交叉测试的训练-测试患者的划分都是相同的。

表 4.2 PSMB-Net 与其它方法在全身 PET/CT 上的比较

模型/指标	Dice	Sensitivity	Precision
W-Net	0.462±0.306	0.565±0.340	0.513±0.346
Two-stream-Net	0.420±0.317	0.429±0.342	0.561±0.369
Co-learn-Net	0.507±0.324	0.515±0.336	0.685±0.355
V-Net	0.465±0.312	0.530±0.338	0.549±0.361
DeepMedic	0.401±0.304	0.453±0.298	0.507±0.378
42% SUV_{max}	0.059±0.136	0.078±0.157	0.091±0.221
50% SUV_{max}	0.045±0.112	0.049±0.111	0.086±0.228
PSMB-Net	0.580±0.287	0.616±0.316	0.688±0.309

如表 4.2，PSMB-Net 在三个指标上都达到了最高。Co-learn-Net 在 Dice 上位于第二名，以 0.073 的差距低于本文方法。接下来的 W-Net 和 V-Net 都达到了接近 0.46 的 Dice，而 Two-stream-Net 和 DeepMedic 分别在 Dice 上达到了 0.420 和 0.401。上述方法的 Sensitivity 和 Precision 大多分布在 0.4 到 0.6 的范围内。两个基于 SUV_{max} 的阈值法的三个指标全部低于 0.1。这些指标说明 PSMB-Net 在全身 PET/CT 图像上是一个更好的方法。

最大值投影（maximum intensity projection, MIP）是通过保留三维图像沿某一方向上的最大值获得，可以从 PET 图像中简单有效地了解患者的整体情况。如图 4.1，展示了两个患者 PET 图像的 MIP 与肿瘤的轮廓，该轮廓包含医生提供的标签与各模型的预测。从中可以看出 PSMB-Net 的优势为：在小肿瘤上更全面的分割和对正常器官的更少误分割。

大部分患者的全身的 PET/CT 图像是例行拍摄的，但是大部分论文仅使用了胸部图片来测试他们模型的性能。为了与这些方法进行更加公平的比较，从全身 PET/CT 图像中截取了胸部来与其它方法再次进行对比。胸部图像是通过肺部的上界与下界之间的区域来选择的。因为胸部图像比全身图像更加简单，而位置门是以一个显著的外观上的边界来划分的，因此在胸部的比较中没有使用位置

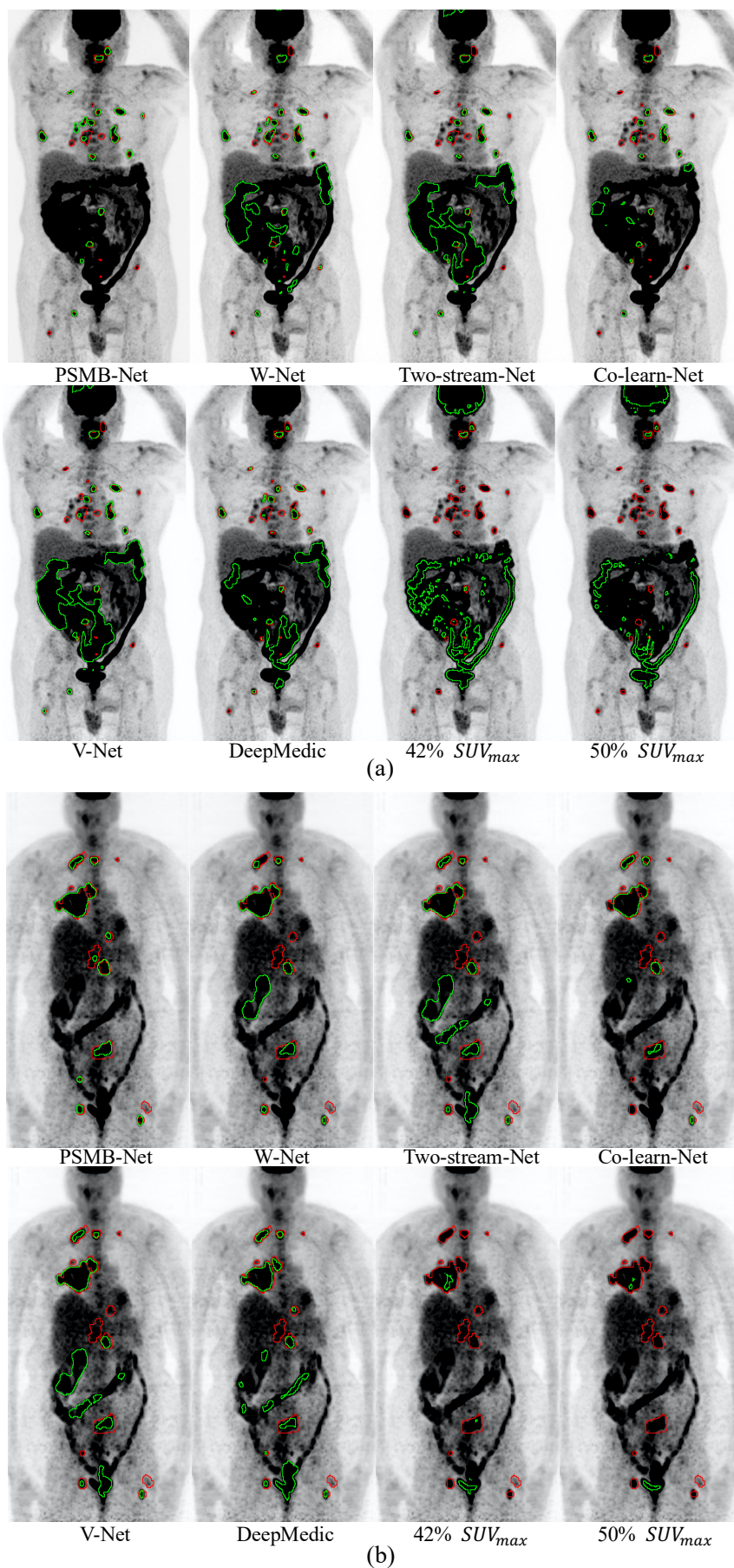


图 4.1 两个患者的 MIP 与肿瘤的边界。红线来自医生的勾画，绿线来自模型输出。

门，即只有一个解码器。

表 4.3 PSMB-Net 与其它方法在胸部 PET/CT 上的比较

模型/指标	Dice	Sensitivity	Precision
W-Net	0.535±0.300	0.676±0.315	0.547±0.331
Two-stream-Net	0.575±0.395	0.643±0.322	0.646±0.331
Co-learn-Net	0.589±0.295	0.660±0.325	0.645±0.317
V-Net	0.572±0.293	0.704±0.307	0.581±0.321
DeepMedic	0.583±0.289	0.577±0.291	0.729±0.335
42% SUV_{max}	0.316±0.270	0.369±0.257	0.498±0.442
50% SUV_{max}	0.272±0.234	0.258±0.205	0.544±0.452
PSMB-Net (无位置门)	0.594±0.283	0.648±0.317	0.673±0.309

在胸部的比较中，如表 4.3，所有方法的指标都相对于各自在全身图像上有提升。PSMB-Net 达到了最高的 Dice，且相比于全身图像提高了 0.014。Co-learn-Net 位于第二名，它相对于自身提高了 0.082。DeepMedic 在 Dice 上提高了 0.182，达到了基于卷积神经网络的方法的最高增量。后续的是 Two-stream-Net，V-Net 和 W-Net，它们的 Dice 分别在各自的基础上提升了 0.155，0.107 和 0.073。42% SUV_{max} 和 50% SUV_{max} 阈值法分别在 Dice 上提高了 0.257 和 0.227。这些指标说明在全身 PET/CT 图像上的肿瘤分割任务相比于胸部图像更困难，使得大部分方法在胸部图像上都取得了更高指标。而且说明 PSMB-Net 在胸部图像上依然适用且有更好的性能。

为了进一步验证 PSMB-Net 核心模块的有效性，本文进行了消融实验，由完整的 PSMB-Net 和 7 种删减版组成，如表 4.4。设置 1：完整的 PSMB-Net；设置 2：PSMB-Net 去掉位置门；设置 3：使用无监督方法产生的位置标签代替手动标记产生的位置标签；设置 4：PSMB-Net 同时去掉位置门和手动标记产生的位置标签；设置 5：PSMB-Net 去掉位置门和位置损失；设置 6：PSMB-Net 去掉多尺度的 CT 特征融合，使用其中最大尺度的特征与 PET 特征进行融合；设置 7：PSMB-Net 去掉类别平衡策略，使用一个固定的系数 20 来放大肿瘤类的梯度；设置 8：PSMB-Net 输出的概率图不使用条件随机场变为二值图，而是使用固定阈值 0.95。

表 4.4 PSMB-Net 在全身 PET/CT 上的消融实验

模型/指标	Dice	Sensitivity	Precision
完整的 PSMB-Net	0.580±0.287	0.616±0.316	0.688±0.309
无位置门	0.546±0.304	0.566±0.324	0.679±0.335
无手动的位置标签	0.539±0.294	0.535±0.323	0.711±0.325
无位置门和手动标签	0.539±0.298	0.549±0.324	0.698±0.329
无位置门和位置监督	0.519±0.305	0.531±0.330	0.662±0.343
无多尺度 CT 特征融合	0.496±0.294	0.583±0.314	0.569±0.346
无平衡系数	0.518±0.321	0.532±0.345	0.684±0.338
无 CRF	0.556±0.259	0.654±0.299	0.614±0.297

如表 4.4, 在设置 2 中把 PSMB-Net 的位置门模块移除后, Dice 下降了 0.036。在设置 2 的基础上, 设置 4 进一步将手动标记产生的位置标签替换为无监督标签后, Dice 又下降了 0.007; 在设置 2 的基础上, 设置 5 直接将位置损失移除, Dice 进一步下降了 0.027。在设置 3 中, 只替换手动标记产生的位置标间后, Dice 比完整的 PSMB-Net 下降了 0.041。在设置 6, 多尺度 CT 特征融合被移除后, Dice 比完整模型下降了 0.084, 这是该实验组中下降最多的设置。在设置 7 中移除肿瘤级的类别平衡策略后, Dice 下降了 0.062, Sensitivity 下降了 0.084, 但是 Precision 只提升了 0.004, 这表明该策略主要通过提升 Sensitivity 来提升整体分割性能。在设置 8 中移除条件随机场后, Sensitivity 提升了 0.038, 但是 Precision 下降了 0.074, 作为综合指标的 Dice 下降了 0.024。这些指标证明这些核心结构在 PSMB-Net 中可以有效提升肿瘤分割的性能。

本文进行了生存分析以验证模型在支持临床决策中的价值。使用的模型是 Cox^[85]和随机生存森林 (RSF)^[86], 如表 4.5。生存时间预测主要使用的是全身肿瘤体积(MTV_{wb}), 该体积分别使用 PSMB-Net 的预测与医生提供的标签计算得到。其输出是 C-index, 是一个生存分析中衡量模型性能的指标^[87]。

在多变量分析中, 两个生存分析模型使用 MTV_{wb} 与其它变量作为自变量, 如全身 SUV_{max} 、临床 TNM 分期、组织学亚型、治疗史、吸烟史、人口统计学变量 (年龄、性别、种族), 以治疗后的总存活时间作为因变量。在生存分析模型在 5 折交叉测试中的每一折中, 使用与肿瘤分割相同的患者分组, 进行训练和测试。在训练阶段, 使用医生提供的 MTV_{wb} 和上述其它变量训练, 在测试阶段, 使用医

生提供的 MTV_{wb} 和模型预测的 MTV_{wb} ，连同其它变量分别测试，然后比较用两种 MTV_{wb} 分别计算出的 C-index。在单变量分析中，C-index 的计算方法与多变量相同，只是训练和测试使用的自变量只有 MTV_{wb} 。

表 4.5 来自 PSMB-Net 与肿瘤标签的 MTV_{wb} 的生存分析对比

模型	MTV_{wb} 来源	单变量分析	多变量分析
Cox	Ground-Truth	0.720±0.019	0.726±0.014
	PSMB-Net	0.697±0.030	0.726±0.013
RSF	Ground-Truth	0.697±0.020	0.738±0.009
	PSMB-Net	0.691±0.024	0.732±0.019

在单变量分析中，使用 PSMB-Net 提供的 MTV_{wb} 在 Cox 和 RSF 中计算出的 C-index 分别比医生提供的 MTV_{wb} 计算出的 C-index 低 0.023 和 0.006。在多变量分析中，使用 PSMB-Net 提供的 MTV_{wb} 在 Cox 中计算出的 C-index 与医生提供的 MTV_{wb} 计算出的 C-index 在小数点 3 位精度内相等，但前者在 RSF 中的 C-index 比后者低 0.006。即使 PSMB-Net 在分割指标上并没有达到非常高的水平，它在生存分析与医生提供的标签相比是可以接受的，这说明本文提出的 PSMB-Net 具有一定的临床决策辅助价值。

4.5 讨论

所有方法在胸部图像的指标都要高于各自在全身图像的指标，这是因为胸部相对于全身来说，其特征是一个子集，这会使模型的学习更有针对性。尤其是在胸部图像中许多高 SUV 区域被先天移除了，唯一的高 SUV 区域是心脏，这使得在全身分割肿瘤相比于胸部来说更加困难。

在 PSMB-Net 的消融实验中，当位置门模块被移除后，Dice 下降了 0.034，这意味着确实有一些肿瘤是用一个解码器无法正确分割的，也证明了位置门这个结构是有效的。

当手动产生的位置标签被无监督方法产生的标签替代后 Dice 有所下降，其原因有两个。第一，对于无监督方法产生的位置标签来说，它包含一小部分切片的相对位置的走势是与整体趋势相反的，比如在图 4.2 (b) 中蓝线的右侧部分。这个问题在提出该方法的论文中同样被提到，其原因是这些切片的内容中有不规

则的姿态。而且所有患者的相对位置走势都不是严格的直线，这其中包含了一定的振动，在 PSMB-Net 的训练过程中可以看作是标签的噪声。PSMB-Net 使用这些带有噪声的标签训练后会产生不准确的位置门信号，导致错误被放大，最终影响肿瘤分割性能。

第二，手动标记产生的位置标签的分布更加集中。如图 4.2 (a)，所有患者的手动产生的相对位置的最大值几乎集中在 1.0 附近，最小值集中在-1.0 附近。但无监督方法产生的相对位置的最大值主要在[0.6, 1.5]这个范围内且变化很大，其最小值主要在[-2.0, -0.2]这个范围内且变化很大。在训练阶段，来自损失函数 \mathcal{L}_{pos} 的梯度被 G_{upper} 和 G_{lower} 加权以训练每一个 BPS-decoder 实现真正地针对于人体的某一部分。在这个要求下， G_{upper} 和 G_{lower} 与人体部位之间的对应关系的稳定性需要被保证。

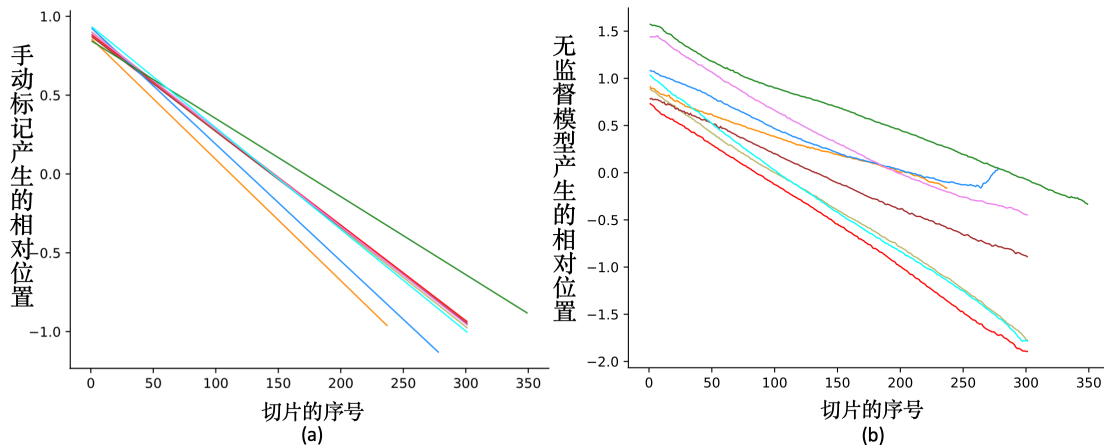


图 4.2 8 个患者由手动标记和无监督模型两种方式产生的相对位置随递增的切片序号变化情况。其中相对位置由 (a) 手动标记和 (b) 无监督模型分别产生。同一患者在两幅图中的颜色相同。

当位置门被移除的同时，手动产生的位置标签也被替换成无监督方法产生的标签（设置 3），导致的 Dice 相比于使用手动产生的标签（设置 4）没有变化。可能的原因是位置标签有两个作用。第一，使 CT 特征图变得位置敏感；第二，用于计算位置门信号。当手动标签被替换为无监督后，CT 特征图与位置门信号都受到了影响，但是 CT 特征图受到的影响程度小于位置门信号受到的影响程度。换句话说，位置门信号对于相对位置的精度更加敏感。

在上述设置（设置 3）的基础上，当位置损失被进一步移除后（设置 5），Dice 也从 0.539 进一步降低了 0.020。这时模型没有使用任何的先验信息。这也证明

了位置损失确实为 CT 特征图提供了额外的信息,可以使其促进模型的肿瘤分割。这可以解释为:对于在本文数据中的低剂量 CT 图像来说,在一个较小的感受野内,一个自身与邻域内的值都高度相似的体素在不同位置都是高度相似的。比如一个在心脏中心的体素与一个在肿瘤中心的体素,它们自身与邻域都具有很高的 SUV,而受到卷积核感受野的限制,很难将二者区分开来。但是在位置损失的监督下,在不同位置的切片内容在特征空间中的距离被放大,导致这些切片的内容中可能出现额外于原图像的对比度,可以被看作额外的位置可分能力。而且,当 CT 特征与 PET 特征结合后,CT 特征的这种能力就扩展到 PET 特征上,导致即使是局部相似的物体,如肿瘤、心脏、脑、肾、膀胱也会被一定程度上赋予这种位置可分性。

当多尺度 CT 特征融合被从完整的 PSMB-Net 中移除后, Dice 从 0.580 降低了 0.084。这说明位置敏感的特征是逐步提取到的,较浅的卷积层输出的 CT 特征图并不具备足够的的能力促进肿瘤分割。

平衡系数用于缓解数据分布的不平衡问题,消融实验表明应用该系数的结果主要是提高了 Sensitivity。其原因是该系数使得整个肿瘤在损失函数中的权重随着肿瘤体素个数的线性增长而缓慢增长,使得小肿瘤与大肿瘤造成的梯度更加接近,导致模型在训练阶段对小肿瘤的关注更加明显。这个系数几乎不会影响对大肿瘤的分割性能,这是因为大肿瘤在测试时跟容易被正确地分割出来。

条件随机场是一个精细化分割结果的有效方法。它通过考虑肿瘤类与非肿瘤类的概率图,同时结合归一化的 PET 原始图像来计算最终二值分割图。当使用条件随机场后, Sensitivity 有小幅下降,但是 Dice 有了一定提升。在使用的 3 个指标中,最关注 Dice 的原因是它是一个综合指标,如果欠分割与过分割率中的任何一个较低,它都会受到影响,而 Sensitivity 和 Precision 只能分别反映二者中的一个。另外,条件随机场也包含一系列的参数,也许存在一个更好的参数集能进一步提升指标。

4.6 本章总结

本章使用 480 个肺癌患者的 PET/CT 图像来验证所提方法。图像分割指标使用了 Dice、Sensitivity 和 Precision。PSMB-Net 在全身图像的比较中,三个指标

均位列第一；在胸部图像的比较中，Dice 位列第一。在全身图像的消融实验中，去掉 PSMB-Net 的任何一个核心模块，Dice 都有所下降，说明这些模块是有效的。在生存分析中，使用 PSMB-Net 计算出的 MTV_{wb} 与生存时间之间的 C-index 与医生的手动勾画结果非常接近。这些实验证明 PSMB-Net 在临床的辅助决策中是有价值的。

5 展望与总结

5.1 展望

在本文所提模型中，使用了 V-Net 型结构作为主干，它包含了残差连接用来缓解梯度消失问题，包含了跳跃连接来保留更多细节信息，并且取得了不错的效果。但是，其它主干网络，如 Dense-Net^[49]，在其它许多任务上比 V-Net 达到了更高的指标，而且我们相信选择主干网络的选择很重要，但是并不把它作为这篇论文的重点，而是将其放在未来的工作中。另外，最近的研究提出了自动机器学习^[88]，其可以自动搜索网络结构和设置超参数。在未来的工作中，会尝试将所提方法与这项技术结合起来，进而获得性能上的突破。

在本文方法中，位置标签依赖手动标记，虽然它有着更高的准确性和稳定性，但是增加了人力负担，因此不是一个完美的解决办法。在前一节提到，对本文方法来说，无监督方法的一个主要缺陷是缺少对不同人的相对位置在同一人体剖面上的相似性约束。一个可能的解决办法是改进其损失函数，以使所有人的位置相对位置分布在更集中的范围内，我们会在未来工作中进行尝试。

本文方法使用的先验信息是 CT 切片在人体的垂直轴上的相对位置，而在另外两个轴上，切片也具有相对位置。除位置外，患者的属性，如年龄、体重、吸烟状态、性别，同样可以作为先验信息，而且将这些文本型信息与图像信息的融合方式也很重要。在未来的工作中会充分考虑先验信息辅助于医学图像处理的可能性，进而做出更多尝试。

5.2 总结

本文立足于从全身 PET/CT 图像中分割肿瘤来辅助医生对癌症患者的诊断和治疗过程。在第一章，首先介绍了研究的意义，然后介绍了现有的 PET/CT 图像分割的几类方法与各自的优缺点，接着列举了该项研究的主要困难和解决思路。由于所提方法以卷积神经网络为基础，在第二章介绍了卷积神经网络的基本概念和核心部分。在第三章，提出了一种以 PSMB-Net 为核心的全身 PET/CT 肿瘤分割方法，以预测 CT 切片在人体中的相对位置为基础，使用门控信号和多尺度特征融合。在第四章通过实验测试所提方法的性能，结果显示 PSMB-Net 在图像指

标和生存分析指标上都取得了不错的成绩。尽管如此，我们认为该方法还有很多改进的空间，因此在第五章提出了未来工作的一些可能性。

参考文献

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: A Cancer Journal for Clinicians, 2021.
- [2] Lee, P., Weerasuriya, D.K., Lavori, P.W., Quon, A., Hara, W., Maxim, P.G., Le, Q.T., Wakelee, H.A., Donington, J.S., Graves, E.E., Loo, B.W.. Metabolic tumor burden predicts for disease progression and death in lung cancer [J]. International Journal of Radiation Oncology*Biology*Physics, 2007, 69(2), 328-333.
- [3] Zhang, C., Liao, C., Penney, B.C., Appelbaum, D.E., Simon, C.A., Pu, Y.. Relationship between overall survival of patients with non-small cell lung cancer and whole-body metabolic tumor burden seen on postsurgical fluorodeoxyglucose PET images [J]. Radiology, 2015, 275(3), 862-869.
- [4] Park, H., Dahlberg, S.E., Lydon, C.A., Araki, T., Hatabu, H., Rabin, M.S., Johnson, B.E., Nishino, M.. M1b disease in the 8th edition of TNM staging of lung cancer: Pattern of single extrathoracic metastasis and clinical outcome [J]. The Oncologist, 2019, 24(8), e749-e754.
- [5] Nestle, U., Kremp, S., Grosu, A.L.. Practical integration of [18F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): The technical basis, ICRU-target volumes, problems, perspectives [J]. Radiotherapy and Oncology, 2006, 81(2), 209-225.
- [6] Schinagl, D.A.X., Vogel, W.V., Hoffmann, A.L., van Dalen, J.A., Oyen, W.J., Kaanders, J.H.A.M.. Comparison of five segmentation tools for 18F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer [J]. International Journal of Radiation Oncology*Biology*Physics, 2007, 69(4), 1282-1289.
- [7] Paulino, A.C., Koshy, M., Howell, R., Schuster, D., Davis, L.W.. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer [J]. International Journal of Radiation Oncology*Biology*Physics, 2005, 61(5), 1385-1392.
- [8] Deniaud-Alexandre, E., Touboul, E., Lerouge, D., Grahek, D., Foulquier, J.N., Petegnief, Y., Grès, B., El Balaa, H., Keraudy, K., Kerrou, K., Montravers, F.,

- Milleron, B., Lebeau, B., Talbot, J.N.. Impact of computed tomography and 18F-deoxyglucose coincidence detection emission tomography image fusion for optimization of conformal radiotherapy in non-small-cell lung cancer [J]. *International Journal of Radiation Oncology*Biophysics*, 2005, 63(5), 1432-1441.
- [9] Nagel, C.C.A., Bosmans, G., Dekker, A.L.A.J., Öllers, M.C., De Ruyscher, D.K.M., Lambin, P., Minken, A.W.H., Lang, N., Schäfers, K.P.. Phased attenuation correction in respiration correlated computed tomography/positron emitted tomography [J]. *Medical Physics*, 2006, 33(6), 1840-1847.
- [10] Erdi, Y., Mawlawi, O., Larson, S., Imbriaco, M., Yeung, H., Finn, R., Humm, J. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding [J]. *Cancer*, 1997, 80, 2505-2509.
- [11] Nestle, U., Kremp, S., Schaefer-Schuler, A., Sebastian-Welsch, C., Hellwig, D., Rube, C., Kirsch, C.M.. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer [J]. *Journal of Nuclear Medicine*, 2005, 46(8), 1342-1348.
- [12] Daisne, J.F., Sibomana, M., Bol, A., Doumont, T., Lonneux, M., Grégoire, V.. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms [J]. *Radiotherapy and Oncology*, 2003, 69(3), 247-250.
- [13] Jentzen, W., Freudenberg, L., Eising, E.G., Heinze, M., Brandau, W., Bockisch, A.. Segmentation of PET volumes by iterative image thresholding [J]. *Journal of Nuclear Medicine*, 2007, 48(1), 108-114.
- [14] Aristophanous, M., Penney, B.C., Martel, M.K., Pelizzari, C.A.. A gaussian mixture model for definition of lung tumor volumes in positron emission tomography [J]. *Medical Physics*, 2007, 34(11), 4223-4235.
- [15] Hatt, M., Cheze le Rest, C., Turzo, A., Roux, C., Visvikis, D.. A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET [J]. *IEEE Transactions on Medical Imaging* 28(6), 881-893.
- [16] Hatt, M., Cheze Le Rest, C., Albarghach, N., Pradier, O., Visvikis, D.. PET functional volume delineation: A robustness and repeatability study [J]. *European Journal of Nuclear Medicine and Molecular Imaging*, 2011, 38(4), 663-672.

- [17] Hofheinz, F., Langner, J., Petr, J., Beuthien-Baumann, B., Steinbach, J., Kotzerke, J., van den Hoff, J.. An automatic method for accurate volume delineation of heterogeneous tumors in PET [J]. *Medical Physics*, 2013, 40(8).
- [18] Hatt, M., Cheze le Rest, C., Descourt, P., Dekker, A., De Ruyscher, D., Oellers, M., Lambin, P., Pradier, O., Visvikis, D.. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications [J]. *International Journal of Radiation Oncology*Biology*Physics*, 2010, 77(1), 301-308.
- [19] Maitre, A.L., Hatt, M., Pradier, O., le Rest, C.C., Visvikis, D.. 53 poster: Impact of the accuracy of tumor functional volume delineation on radiotherapy treatment planning [J]. *Physics in Medicine and Biology*, 2012, 57(17), 5381-5397.
- [20] Figueiredo, B., Merlin, T., Henri, D.C.G., Hatt, M., Vimont, D., Fernandez, P., Lamare, F.. Potential of [F-18]-fluoromisonidazole positron-emission tomography for radiotherapy planning in head and neck squamous cell carcinomas [J]. *Strahlentherapie und Onkologie*, 2013, 189, 1015-1019.
- [21] Pham, D., Xu, C., Prince, J.. Current methods in medical image segmentation [J]. *Annual Review of Biomedical Engineering*, 2000, 2(2000), 315-337.
- [22] Ma, Z., Tavares, J.M.R.S., Jorge, R.N.. Segmentation of structures in medical images: review and a new computational framework [C]. *Proceedings of 8th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering*, 2008.
- [23] Kerhet, A., Small, C., Quon, H., Riauka, T., Greiner, R., Mcewan, A., Roa, W.. Segmentation of lung tumours in positron emission tomography scans: A machine learning approach [C]. *Proceedings of the 12th Conference on Artificial Intelligence in Medicine*, 2009, 146-155.
- [24] Sharif, M.S., Abbod, M., Amira, A., Zaidi, H.. Artificial neural network-based system for PET volume segmentation [J]. *International journal of biomedical imaging*, 2010.
- [25] Yu, H., Caldwell, C., Mah, K., Poon, I., Balogh, J., MacKenzie, R., Khaouam, N., Tirona, R.. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images [J]. *International Journal of Radiation Oncology*Biology*Physics*, 2009, 75(2), 618-625.

- [26] Belhassen, S., Zaidi, H.. A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in PET [J]. *Medical physics*, 2010, 37(3), 1309-1324.
- [27] Yang, F., Grigsby, P.W.. Delineation of FDG-PET tumors from heterogeneous background using spectral clustering [J]. *European Journal of Radiology*, 2012, 81(11), 3535-3541.
- [28] Foster, B., Luna, B., Dey, B., Bishai, W., Jain, S., Xu, Z., Mollura, D.. Robust segmentation and accurate target definition for positron emission tomography images using Affinity Propagation [C]. *IEEE 10th International Symposium on Biomedical Imaging*, 2013, 1461-1464.
- [29] Day, E., Betler, J., Parda, D., Reitz, B., Kirichenko, A., Mohammadi, S., Miften, M.. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients [J]. *Medical Physics*, 2009, 36(10), 4349-4358.
- [30] Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts [C]. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, 1, 377-384.
- [31] Bağci, U., Yao, J., Caban, J., Turkbey, E., Aras, O., Mollura, D.J.. A graph-theoretic approach for segmentation of PET images [C]. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, 8479-8482.
- [32] Wechsler, H., Kidode, M.. A random walk procedure for texture discrimination [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, 1(3), 272-280.
- [33] Hsu, C.Y., Liu, C.Y., Chen, C.M.. Automatic segmentation of liver PET images [J]. *Computerized Medical Imaging and Graphics: The Official Journal of The Computerized Medical Imaging Society*, 2008, 32(7), 601-610.
- [34] Li, H., Thorstad, W.L., Biehl, K.J., Laforest, R., Su, Y., Shoghi, K.I., Donnelly, E.D., Low, D.A., Lu, W.. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours [J]. *Medical Physics*, 2008, 35(8), 3711-3721.
- [35] Geets, X., Lee, J.A., Bol, A., Lonneux, M., Grégoire, V.. A gradient-based method for segmenting FDG-PET images: Methodology and validation [J]. *European Journal of Nuclear Medicine and Molecular Imaging*, 2007, 34(9), 1427-1438.

- [36] Werner-Wasik, M., Nelson, A.D., Choi, W., Arai, Y., Faulhaber, P.F., Kang, P., Almeida, F.D., Xiao, Y., Ohri, N., Brockway, K.D., Piper, J.W., Nelson, A.S.. What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom [J]. *International Journal of Radiation Oncology*Biology*Physics*, 2012, 82(3), 1164-1171.
- [37] El Naqa, I., Yang, D., Apte, A., Khullar, D., Mutic, S., Zheng, J., Bradley, J., Grigsby, P., Deasy, J.. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning [J]. *Medical physics*, 2007, 34(12), 4738-4749.
- [38] Khatchadourian, S., Lebonvallet, S., Herbin, M., Liehn, J.C., Ruan, S.. Tumor segmentation from PET/CT images using level sets method [C]. *IFAC Proceedings Volumes*, 2006, 39(18), 255-260.
- [39] Markel, D., Zaidi, H., El Naqa, I.. Novel multimodality segmentation using level sets and Jensen-Rényi divergence [J]. *Medical Physics*, 2013, 40(12).
- [40] Yu, H., Caldwell, C., Mah, K., Mozeg, D.. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning [J]. *IEEE Transactions on Medical Imaging*, 2009, 28, 374-383.
- [41] Watabe, H., Markel, D., Caldwell, C., Alasti, H., Soliman, H., Ung, Y., Lee, J., Sun, A. Automatic segmentation of lung carcinoma using 3D texture features in 18-FDG PET/CT [J]. *International Journal of Molecular Imaging*, 2013.
- [42] Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X.. Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method [C]. *International Conference on Information Processing in Medical Imaging*, 2011, 245-256.
- [43] Song, Q., Bai, J., Han, D., Bhatia, S., Sun, W., Roche, W., Bayouth, J.E., Bu-atti, J.M., Wu, X.. Optimal co-segmentation of tumor in PET-CT images with context information [J]. *IEEE Transactions on Medical Imaging*, 2013, 32(9), 1685-1697.
- [44] Cui, H., Wang, X., Lin, W., Zhou, J., Eberl, S., Feng, D., Fulham, M.. Primary lung tumor segmentation from PET-CT volumes with spatial-topological constraint [J]. *International Journal of Computer Assisted Radiology and Surgery*, 2016, 11(1), 19-29.
- [45] Hanzouli-Ben Salah, H., Lapuyade-Lahorgue, J., Bert, J., Benoit, D., Lambin, P., Van Baardwijk, A., Monfrini, E., Pieczynski, W., Visvikis, D., Hatt, M.. A frame-

- work based on hidden Markov trees for multimodal PET/CT image co-segmentation [J]. *Medical Physics*, 2017, 44(11), 5835-5848.
- [46] Wang, X., Ballangan, C., Cui, H., Fulham, M., Eberl, S., Yin, Y., Feng, D.. Lung tumor delineation based on novel tumor-background likelihood models in PET-CT images [J]. *IEEE Transactions on Nuclear Science*, 2014, 61(1), 218-224.
- [47] Potesil, V., Huang, X., Zhou, X.. Automated tumor delineation using joint PET/CT information [C]. *Medical Imaging 2007: Computer-Aided Diagnosis*, 2007, 6514, 928-935.
- [48] Lian, C., Ruan, S., Denoeux, T., Guo, Y., Vera, P.. Accurate Tumor Segmentation In FDG-PET Images With Guidance Of Complementary CT Images [C]. *IEEE International Conference on Image Processing*, 2017, 4447-4451.
- [49] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.. Densely connected convolutional networks [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 2261-2269.
- [50] Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.. Feature pyramid networks for object detection [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 936-944.
- [51] Girshick, R.. Fast R-CNN [C]. *IEEE International Conference on Computer Vision*, 2015, 1440-1448.
- [52] Mehta, S., Rastegari, M., Caspi, A., Shapiro, L.G., Hajishirzi, H.. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation [C]. *European Conference on Computer Vision*, 2018, 561-580.
- [53] He, K., Zhang, X., Ren, S., Sun, J.. Deep residual learning for image recognition [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 770-778.
- [54] Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.. Res2Net: A new multi-scale backbone architecture [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2), 652-662.
- [55] Huang, B., Chen, Z., Wu, P.M., Ye, Y., Feng, S.T., Wong, C.Y.O., Zheng, L., Liu, Y., Wang, T., Li, Q., Huang, B.. Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: A dual-center study [J]. *Contrast Media & Molecular Imaging*, 2018.
- [56] Xu, L., Tetteh, G., Lipkova, J., Zhao, Y., Li, H., Christ, P., Piraud, M., Buck, A., Shi, K., Menze, B.H.. 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.
- [57] Milletari, F., Navab, N., Ahmadi, S.A.. V-Net: Fully convolutional neural networks for volumetric medical image segmentation [C]. *International Conference on 3D Vision*, 2016, 565-571.
- [58] Jin, D., Guo, D., Ho, T.Y., Harrison, A.P., Xiao, J., Tseng, C. k., Lu, L.. Accurate esophageal gross tumor volume segmentation in PET/CT using two-stream chained 3D deep network fusion [C]. *Medical Image Computing and Computer Assisted Intervention*, 2019, 11765, 182-191.
- [59] Zhong, Z., Kim, Y., Plichta, K., Allen, B.G., Zhou, L., Buatti, J., Wu, X.. Simultaneous co-segmentation of tumors in PET-CT images using deep fully convolutional networks [J]. *Medical Physics*, 2019, 46(2), 619-633.
- [60] Li, L., Zhao, X., Lu, W., Tan, S.. Deep learning for variational multi-modality tumor segmentation in PET/CT [J]. *Neurocomputing*, 2019, 392, 277-295.
- [61] Zhao, X., Li, L., Lu, W., Tan, S.. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network [J]. *Physics in Medicine & Biology*, 2018, 64(1).
- [62] Kumar, A., Fulham, M., Feng, D., Kim, J.. Co-learning feature fusion maps from PET-CT images of lung cancer [J]. *IEEE Transactions on Medical Imaging*, 2020, 39(1), 204-217.
- [63] 邱锡鹏.神经网络与深度学习[M].机械工业出版社: 2020.
- [64] 魏秀参.解析深度学习——卷积神经网络原理视觉实践[M].电子工业出版社: 2018.
- [65] Goodfellow, I., Bengio, Y., Courville, A.. *Deep Learning* [M]. MIT Press: 2016.
- [66] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.. Gradient-based learning applied to document recognition [C]. *Proceedings of the IEEE*, 1998, 86(11), 2278-2324.
- [67] He, K., Zhang, X., Ren, S., Sun, J.. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]. *European Conference on Computer Vision*, 2014, 346-361.
- [68] Han, J., Moraga, C.. The influence of the sigmoid function parameters on the speed of backpropagation learning [C]. *Proceedings of the International Workshop on Artificial Neural Networks: From Natural to Artificial Neural Computation*, 1995, 195-201.

- [69]Glorot, X., Bordes, A., Bengio, Y.. Deep sparse rectifier neural networks [C]. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, 15, 315-323.
- [70]Clevert, D.A., Unterthiner, T., Hochreiter, S.. Fast and accurate deep network learning by exponential linear units (ELUs), 2016.
- [71]He, K., Zhang, X., Ren, S., Sun, J.. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification [C]. IEEE International Conference on Computer Vision, 2015, 1026-1034.
- [72]<https://image-net.org> [Z].
- [73]Glorot, X., Bengio, Y.. Understanding the difficulty of training deep feedforward neural networks [J]. Journal of Machine Learning Research-Proceedings Track, 2010, 9, 249-256.
- [74]Smith, L.N.. Cyclical learning rates for training neural networks [C]. IEEE Winter Conference on Applications of Computer Vision, 2017, 464-472.
- [75]Loshchilov, I., Hutter, F.. Stochastic gradient descent with warm restarts [C]. International Conference on Learning Representations, 2017.
- [76]Li, Z., Zhang, S., Zhang, J., Huang, K., Wang, Y., Yu, Y.. MVP-Net: Multi-view FPN with position-aware attention for deep universal lesion detection [C]. Medical Image Computing and Computer Assisted Intervention, 2019, 13-21.
- [77]Kirillov, A., Girshick, R., He, K., Dollár, P.. Panoptic feature pyramid networks [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 6392-6401.
- [78]Zhu, Z., Jin, D., Yan, K., Ho, T. Y., Ye, X., Guo, D., Chao, C. H., Xiao, J., Yuille, A., Lu, L.. Lymph node gross tumor volume detection and segmentation via distance-based gating using 3D CT/PET imaging in radiotherapy [C]. Medical Image Computing and Computer Assisted Intervention, 2020, 753-762
- [79]Krähenbühl, P., Koltun, V.. Efficient inference in fully connected CRFs with gaussian edge potentials [J]. Advances in Neural Information Processing Systems, 2012, 24, 109-117.
- [80]<https://github.com/ashnilkumar/colearn> [Z].
- [81]Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Medical Image Analysis, 2017, 36, 61-78.

- [82] Foster, B., Bagci, U., Mansoor, A., Xu, Z., Mollura, D.J.. A review on segmentation of positron emission tomography images [J]. *Computers in biology and medicine*, 2014, 50, 76-96.
- [83] <https://github.com/deepmedic/deepmedic> [Z].
- [84] Yan, K., Lu, L., Summers, R.M.. Unsupervised body part regression via spatially self-ordering convolutional neural networks [C]. *IEEE 15th International Symposium on Biomedical Imaging*, 2018, 1022-1025.
- [85] Cox, D.R.. Regression models and life tables [J]. *Journal of the Royal Statistical Society*, 1972, 34(2), 187-220.
- [86] Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.. Random survival forests [J]. *The Annals of Applied Statistics*, 2008, 2(3), 841-860.
- [87] Raykar, V.C., Steck, H., Krishnapuram, B., Dehing-Oberije, C., Lambin, P.. On ranking in survival analysis: Bounds on the concordance index [C]. *Advances in Neural Information Processing Systems*, 2007, 20, 1209-1216.
- [88] Yao, Q., Wang, M., Escalante, H.J., Guyon, I., Hu, Y., Li, Y., Tu, W., Yang, Q., Yu, Y.. Taking human out of learning applications: A survey on automated machine learning. *CoRR abs/1810.13306*, 2018.

致谢

硕士这三年即将结束，在这段时间里我收获良多。在此，我感谢帮助过我的老师、父母和同学。

首先，我要感谢我的导师，何江萍老师。何老师对我严格要求，让我养成了良好的科研习惯。他对待学术的严谨态度让我肃然起敬，他对前沿技术的探索精神总给我以鼓舞。每当我有难以解决的问题时，都能向他请教，让我可以顺利完成实验和论文。

然后我要感谢韩金仓老师、丁晓阳老师。他们为实验室的建设贡献了很大力量，为我的科研提供了重要支持。他们丰富的阅历为我提供了宝贵的经验，使我的眼界更加开阔。

其次，我要感谢李强老师的电子商务实验室，让我使用了性能优秀的计算机，大大加快了我实验的进程，为我按时完成论文提供了保障。

最后，我要感谢我的父母、实验室同学。我的父母承担了我上学期间所有的费用，对我的生活非常关心，并给予鼓励和开导，是我坚强的后盾。感谢父母！实验室的高洪萍同学在代码和数据上为我打下了重要的基础，刘天宇同学为我提供了思路和学习上的交流。感谢他们！

感谢所有评阅我论文的老师，谢谢你们抽出宝贵的时间对我的论文进行指导。

硕士期间科研项目与成果

参与的科研项目：

- 1 国家自然科学基金：“三维肝脏 CT 图像中肝脏肿瘤的计算机辅助诊断方法研究”（项目批准号：61661024）
- 2 甘肃省自然科学基金：“PET/CT 图像中肿瘤组织类型自动判别方法研究”（项目批准号：20JR5RA200）

取得的研究成果：

Kun Wang, Maggie Chung, Michael Wang, Yan Ma, Qiang Li, Xiaoyang Ding, Jiangping He, Yonglin Pu: Metabolic Tumor Segmentation from Whole-body PET/CT Images Using a Position Supervised Multi-branch Neural Network (Submitted to IEEE Transactions on Medical Imaging, SCI 一区)